THE UNIVERSITY
of EDINBURGH

Ligand Discovery for Protein-Protein Interaction Targets Using ^{19}F NMR-Based Screening of Novel Peptide and Fragment Libraries

Thesis submitted in accordance with the requirement of The University of Edinburgh for the Degree of Doctor of Philosophy

By

Ian Spink

Supervised by

Prof. Manfred Auer

School of Biological Sciences

College of Science and Engineering

February 2018

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Signed

(Ian Spink)

Date 24/08/2018

Acknowledgements

First and foremost, I would like to thank my principal supervisor Professor Manfred Auer for his guidance, encouragement and support throughout my time as a PhD student. I would also like to extend my thanks to the other members of my thesis committee, Dr Gary Blakely and Professor Paul Barlow, for their support and willingness to help when required. Also, thanks must be given to the MRC for funding for the duration of my PhD studentship. I am especially indebted to all the collaborators on this work, here, and from the University of Bayreuth. Stefan, and all the members of the Department of Biopolymers, for making my stay in Germany a wonderful experience.

I am grateful to all the technical and service staff in the School of Biology and School of Chemistry. A special thanks goes to the members of the NMR suite in the School of Chemistry, Juraj Bella and Lorna Murray, for their trusted guidance, patience and expertise in NMR spectroscopy. A big thank you to all members of the SynthSys department and the whole of the Waddington building for their kindness and friendship throughout my time spent there.

Now, a list of thank-yous to rival an Oscars acceptance speech: to all the members of the Auer group, past and present, I must extend my sincerest gratitude for everything you have done for me. Nhan for his invaluable technical knowledge and friendship, Steve for all of the computational work, advice and company. To Joanna, for her keen insight and brilliant practical knowledge in the field of biology. To e-Pi for the company in the lab, all the laughs and so many kilos of cheese I can't even count. To the other members of the group, Stefan, Dave, Geordie, Gemma, Olivier, Fethers, I am especially grateful to all without whom I would not have got this far. To all members of Edinburgh Jitsu Club (Neil, Andy and Dave, in particular), I give thanks for letting me release some steam and providing an essential distraction during my time (oh, and for all the beers of course!).

To all members of my family, parents and brothers, I cannot begin to describe just how much your kindness has meant over my time as a PhD student, I will never forget it.

Finally, and most importantly, I must thank my loving and supportive wife, Hannah. Nobody has been more important to me over the course of this PhD and I would not have made it this far without you. You're unending kindness, support and encouragement has been an inspiration and I will be eternally grateful.

List of Abbreviations

3-FABS	Three Fluorine Atoms for Biochemical Screening
AAA	ATPase Associated with Diverse Cellular Activities
ACN	Acetonitrile
AQ	Acquisition time
ATAD2	ATPase AAA Domain Containing 2
ATP	Adenosine triphosphate
AU	Abitrary units
Boc	*tert*-Butyloxycarbonyl
BRD	Bromodomain
CFL	Clustered Fragment Libray
C-FTL	C-terminal Fluorine Tripeptide Library
CONA	Confocal Nanoscanning
COPAS	Complex Object Parametric Analyser and Sorter
CPMG	Carr-Purcell-Meiboom-Gill
CSA	Chemical Shift Anisotropy
CSP	Chemical Shift Perturbation
CTB	Chemical and Translational Biology
Cy5	Cyanine-5
d	Doublet
Da	Dalton
DAD	Diode Array Detector
DCM	Dichloromethane
dCSP	Differential Chemical Shift Perturbation
DDFL	Directed Diversity Fragment Library
DIPEA	N,N-Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
DOA	Dioxaoctanoic acid
DTT	Dithiothreitol
ESI	Electrospray Ionisation

FAXS	Fluorine Anisotropy and Exchange for Screening
FBA	3-Fluoro-2-hydroxy benzoic acid
FBDD	Fragment Based Drug Design
FBS	Fragment Based Screening
FCS	Fluorescence Correlation Spectroscopy
FDL	Fluorine Dipeptide Library
FLD	Fluorescent detector
FMOC	9-Fluorenylmethyloxycarbonyl
GSK	GlaxoSmithKline
HA	Heavy Atoms
HATU	O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High Performance Liquid Chromatography
hr	Hour(s)
HSA	Human Serum Albumin
HSQC	Heteronuclear Single Quantum Coherence
HTS	High Throughput Screening
Hz	Hertz
IAP	Inhibitor of Apoptosis
ILOE	Interligand Overhauser Effect
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUPAC	International Union of Pure and Applied Chemistry
K_D	Dissociation constant
LCMS	Liquid Chromatography Mass Spectrometry
LE	Ligand Efficiency
m	multiplet
m/z	Mass to charge ratio
MeOH	Methanol
MES	2-(N-morpholino)ethanesulfonic acid
MHz	Megahertz
min	Minute(s)
MS	Mass spectrometry
MW	Molecular Weight

MWCO	Molecular Weight Cut Off
N-FTL	N-terminal Fluorine Tripeptide Library
NHS	N-hydroxy succinimide
Ni-NTA	Nickel nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NROT	Number of Rotational Bonds
NS	Number of Scans
Nus	N-utilisation substance
OBOC	One Bead One Compound
OBS	On Bead Screening
OD	Optical Density
OPPG	On Protein Peptide Growing
PAIN	Pan Assay Interference
PBD	Protein Data Bank
PBS	Phosphate Buffered Saline
PPI	Protein-Protein Interaction
ppm	Parts per Million
PRA	Propargyl glycine
PS/PS	Post Synthesis Post Screening
PSA	Polar Surface Area
PTM	Post Translational Modification
q	quarter
QC	Quality control
r.t.	Room Temperature
RAFL	Recognised Active Fragment Library
RED	Rapid Equilibrium Device
RNA	Ribonucleic acid
RNAP	Ribonucleic acid Polymerase
rpm	Rotations per Minute
s	singlet
SAR	Structure Activity Relationship

SCAL2	Safety Catch Acid Labile linker
SDF	Structure Data File
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis
SFO1	Spectrometer Operating Frequency
SGC	Structural Genomics Consortium, Oxford
SL	Spin Lock
SM	Small Molecule
SMILES	Simplified Molecular Line Entry System
SPE	Solid-phase Extractor
SPPS	Solid-phase Peptide Synthesis
STD	Saturation Transfer Difference
SW	Spectral width
t	triplet
TAMRA	Carboxytetramethylrhodamine
TCEP	Tris(2-carboxyethyl)phosphine
TD	Time domain
TFA	Trifluoroacetic acid
TG	TentaGel
TG-RAM	TentaGel Rink Amide Resin
TIS	Triisopropyl Silane
TLC	Thin Layer Chromatography
TMR	Tetramethylrhodamine
TMSBr	Trimethylsilyl bromide
TNBS	Trinitrobenzene sulfonic acid
tR	Retention time
Ub	Ubiquitin
UbE2L3	Ubiquitin Conjugating Enzyme E2 L3
UPLC	Ultra Performance Liquid Chromatography
UPS	Ubiquitin Proteasome System
USR	Ultrafast Shape Recognition
USRCAT	Ultrafast Shape Recognition with CREDO Atom Types
UV-Vis	Ultraviolet-Visible Spectroscopy

Abstract

The main aim of this thesis was to discover and design new ligands for difficult, under-explored and clinically relevant protein targets. A number of protein-protein interaction complexes (PPIs) are introduced as the target focus for the methods employed and developed herein. This thesis is separated into two sections to independently address both peptides and small molecules as screening agents. The project examines both approaches through comprehensive library design strategies and screening by NMR spectroscopic methods.

ATAD2 is the first PPI investigated and was expressed and purified in good yield and was also isotopically labelled with Nitrogen-15 for enhanced sensitivity and orthogonal ligand and protein-observed NMR methods. A known pentapeptide was synthesised by solid-phase peptide synthesis (SPPS) using Fmoc chemistry for target validation and tool compound development. A one-bead one-compound (OBOC) tripeptide library was synthesised by SPPS in good yield and purity, determined using single-bead labelling techniques with a fluorescent dye (TMR) and HPLC analysis. This library contained 3072 unique tripeptides with 12 central non-natural, lysine derivatives flanked by 16 natural *L* amino acids. The library screening technique was based on using a fluorescently labelled protein and Confocal Nanoscanning to detect binding. However, fluorescent labelling of ATAD2 was unsuccessful due to difficult protein handling conditions, therefore this library was not screened. The advent of small molecule, high affinity inhibitors of this target protein generated by GSK shifted focus to a different PPI target, the ubiquitin conjugating enzyme, UbE2L3.

A novel "on-protein peptide building" approach was introduced with the aim of screening a library of fluorinated dipeptides and extending the most potent *via* the 'N' and 'C' terminus to increase the affinity. A proof-of-concept tetrapeptide to survivin was synthesised by SPPS by incorporation of a non-natural, fluorinated amino acid in the known tetrapeptide sequence. This fluorinated derivative showed target binding activity by ^{19}F NMR spectroscopy. The tripeptide and dipeptide truncates were synthesised by SPPS and binding was still observable by ^{19}F NMR. This method was extended to screening a library of synthesised fluorinated dipeptides by ^{19}F NMR against UbE2L3. A single dipeptide was identified with low affinity and the dipeptide was extended C and N terminally by SPPS to increase affinity. However, there were no tripeptides identified for this protein using this method. The proof of concept

tetrapeptide was a success, therefore further protein targets are required to conclusively assess the viability of the approach.

Fragment based screening is then introduced as a second approach to novel ligand discovery. Coupled with cheminformatics analysis and *in silico* library design, we created an in-house fluorinated fragment library consisting of 109 fluorinated fragments using three parallel methods. Compounds were purchased and quality checked by LCMS, HPLC and ^{19}F-NMR. These fragment libraries were screened in a ^{19}F NMR assay against the UbE2L3 and NusE/NusB protein targets. In a primary mixture screen, two fragment hits were identified against the NusE/NusB PPI and there were no fragment hits identified against the UbE2L3 protein. The two fragments against NusE/NusB were validated using orthogonal ligand-binding NMR methods. A mini-series, consisting of six commercially available analogues, were purchased and two fragment analogues showed increased affinity and were active against *E. coli* in a bacterial inhibition assay. The dissociation constants of the six active compounds were determined by ^{15}N-HSQC NMR titration experiments and shown to be in 100 – 500 μM range. The binding sites of each compound were also determined by ^{15}N-HSQC chemical shift mapping. These fragment hits represent a novel chemical scaffold identified against the NusE/NusB PPI and demonstrate the potential druggability of this new, complex target.

The use of fluorine as a sensor for binding detection is evaluated by incorporating into both peptides and fragments. Through the use of novel library design strategies, a campaign to discover novel ligands of difficult protein targets is presented.

Lay Summary

Drug discovery is the process by which new candidate medicines are discovered and developed. Typically, drugs work by interacting with a biological protein inside cells and moderating their function. Crucial to the drug discovery process is to discover new proteins that can be targeted with drugs, and one class of increasingly popular targets for drug intervention are binding partners of protein-protein interactions (PPIs). There are about 200,000 PPIs inside cells at any point of time and they are essential for controlling multiple cellular functions. Any discrepancies in the function of PPIs can lead to diseases, including: cancer, neurodegenerative diseases, infection and many more. Three main types of drugs are used on the market. Proteins, peptides and their structural elements, and small molecules.

One of the most common ways to discover a new drug is to screen many thousands of compounds (called a library) simultaneously against the target of interest and determine which one has the largest effect.

In this project, the aim was to develop novel, more efficient ways, of designing libraries of compounds, of both peptides and small molecules, which would maximise the chances of finding a potential drug lead. The new libraries were then evaluated for potential interaction with a number of different PPI targets using a sensitive physical technique known as Nuclear Magnetic Resonance spectroscopy (NMR). This technique can measure compounds that have a very weak interaction with the target of interest, therefore further maximising the chances of finding a potential starting point for drug development. The chosen PPI targets for this project have previously been classed as difficult to find chemical starting points, therefore the combination of efficient library design and a sensitive detection method, was applied to maximise the chances of finding a molecule which would bind to the target. One of the PPIs introduced as target in this thesis is important for the growth of *E. coli* bacteria, called NusE/NusB. Finding a potential drug compound for this will disrupt the growth and prevent bacterial infection. From our novel libraries two small molecules were identified which bind to this important regulator of bacterial growth, NusE/NusB. These small molecules had affected the growth of *E. coli* bacteria, and therefore represent starting points for further chemical modification and development into more potent inhibitors.

Table of Contents

Chapter 1

Introduction

This thesis is focussed on the discovery of new ligands for difficult, under-explored protein targets by strategic development and application of methods in library design and NMR-based screening approaches. This chapter begins with a general project overview and aims of the research. Then a brief introduction to the supporting background of the drug discovery process and a literature review of the various aspects of target selection and screening technologies.

1.0 General Project Outline

Drug discovery is the process by which new candidate medicines are discovered and developed. Discovery of new and improved therapeutics involves an extensive pipeline from target validation and hit discovery to clinical trials. Target validation is crucial to find the right starting points for the initial hit and lead discovery phase. The most challenging target class for drug intervention are protein-protein interactions (PPIs). The main aim of this work is to discover and design new ligands using novel methods of library design and to evaluate these libraries against difficult, under-explored and clinically relevant PPI targets. This is achieved through a combination of two approaches: peptides and fragments. Both approaches involve the design of fluorinated compound libraries and subsequent screening of protein targets by ligand-observed ^{19}Fluorine NMR.

The first approach describes the synthesis of novel, target focussed 'One Bead One Compound' (OBOC) peptidomimetic libraries that can be used for screening of novel binders. The aim of this work is to discover novel, non-natural tripeptides and peptidomimetics designed to bind to the epigenetic, cancer target ATAD2 (ATPase AAA Domain Containing 2). Firstly, the synthesis of a known peptide ligand to ATAD2 is described and NMR methods for binding detection are reported. For ligand discovery, a focussed, one-bead one-compound peptide library is synthesised by a solid-phase combinatorial split-and-mix strategy using Fmoc chemistry. Methods for screening and binding detection are described in this section.

An alternative peptidic approach to discover and build novel peptide ligands, named 'On Protein Peptide Growing', using chemical libraries of fluorinated peptides is also described in chapter 2. Exploiting the high sensitivity of NMR-based detection methods, a fluorinated tetrapeptide, to the cancer target survivin, is used to assess the feasibility of the approach. The method involves the synthesis of an initial fluorinated dipeptide of the N-terminus of the known tetrapeptide. This is then extended by iterative 'C' terminal addition of amino acids, into a novel peptide building approach. Binding of the fluorinated dipeptide is detected by ^{19}F NMR and increasing binding affinity is observed after C-terminal extension towards the final tetrapeptide. The approach was then applied to a validated, difficult target, the ubiquitin conjugating enzyme, UbE2L3. An initial dipeptide library consisting of a fluorinated phenylalanine and a natural *L* amino acid were screened for binding activity. The fluorine chemical shift of the fluorinated phenylalanine present in peptides with different amino acid

composition and/or length is distinctly different to measure in a screening approach. Multiple libraries of C- and N-terminal extended tripeptides were synthesised, to discover a new ligand for this target.

The second approach presents fluorinated fragments as compounds for screening campaigns. The aim of this work was to design, collect and evaluate the fragment libraries to discover new, fragment hit compounds against orphan, clinically relevant PPI drug targets, including: UbE2L3 and NusE/NusB. The design of fragment libraries was critical to the success of the approach and a novel method for fragment library design was introduced. A concise description of practical aspects to screening by NMR and optimisation of assay parameters for a known fluorinated fragment-sized ligand for HSA is included.

The in-house, designed fragment libraries were screened against the next-generation bacterial PPI target, NusE/NusB. Hit fragments identified by ^{19}F NMR CSP were validated by secondary NMR methods, including T_2 relaxation and ^{15}N-HSQC, and a propriety version of biochemical assay, quantitative microdialysis (qμD). Also, structure-guided approaches to hit fragment development were employed, including a SAR by catalogue approach. The binding sites of hit compounds and analogues were mapped using the assigned, backbone amide residues of ^{15}N-HSQC chemical shift perturbation data. The hit fragments and analogues were also tested for bactericidal activity on *E. coli* cells in a plate-based bacterial growth inhibition assay.

The objective of this thesis is in the design, collection and analysis of fluorine-based chemical libraries and to evaluate them against clinically relevant protein targets. This thesis is organised as follows:

Chapter 1 – Begins with a general project outline and aims of the research. Then a brief introduction to drug discovery and a background to the theory of peptide and fragment-based screening methods.

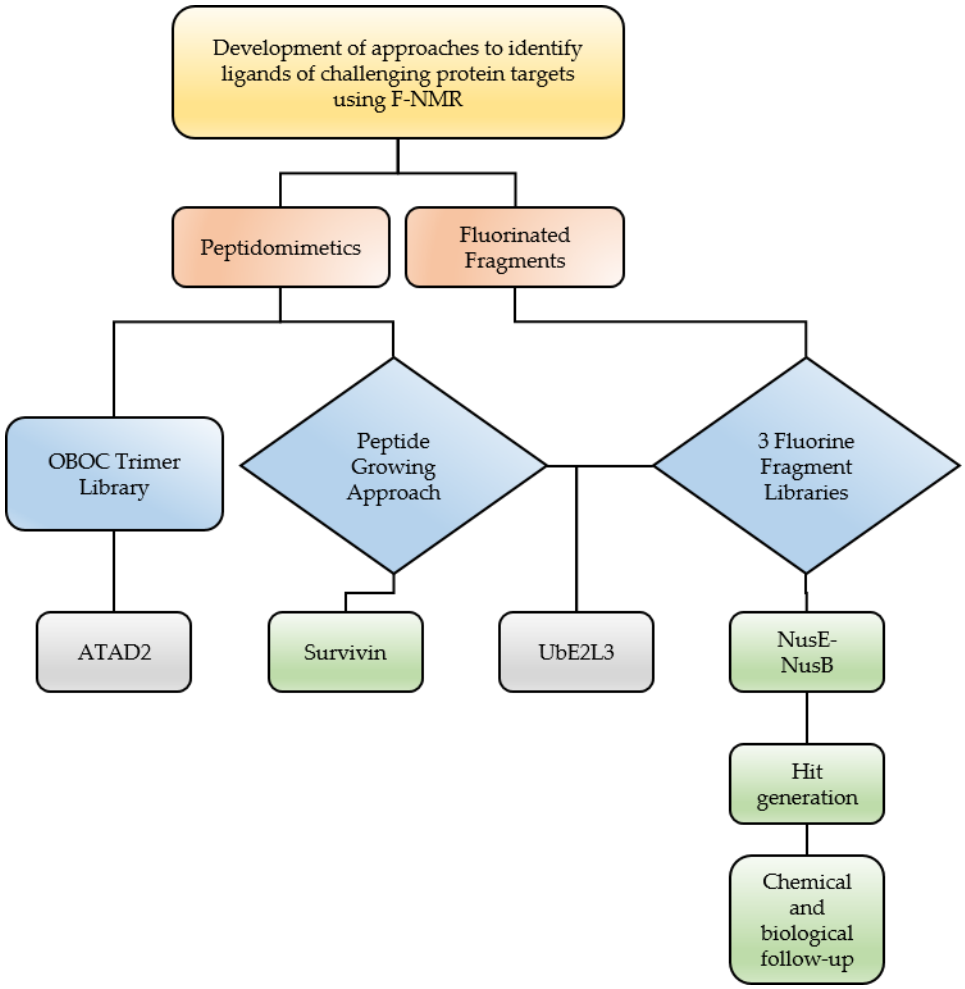
Chapter 2 – Introduces peptides as therapeutic agents and methods of peptide screening. The results section describes the synthesis of peptides for tool compound development for ATAD2 and survivin. The various methods used to design, synthesise and evaluate peptide libraries into a novel "peptide building" approach are also described in this chapter.

Chapter 3 - A comprehensive review of commercial fragment libraries and a novel approach to fragment library design is discussed in chapter 3.

Chapter 4 – Establishment of the designed, in-house CTB fluorinated fragment libraries is described and evaluation against the UbE2L3 protein target is presented.

Chapter 5 – Evaluation and screening of the fragment libraries against a bacterial PPI target NusE-NusB is described. Chemical and biological validation and follow-up of hit compounds is also presented.

Chapter 6 – Concluding remarks and future outlook.



This workflow illustrates the methods employed in this thesis and their application to a range of protein targets.

1.1 Introduction and background

1.1.1 What is drug discovery?

It is an axiom to suggest that the discovery of new drugs is paramount to the continued existence of the human race. For centuries, medicines derived from natural sources – either discovered by mistake or analytically through testing – have been used to cure various ailments and diseases. One of the most significant and earliest examples (*ca.* 1552) of drug discovery is morphine, isolated as the first active alkaloid from the opium poppy plant in 1824 (Figure 1.1).[1]

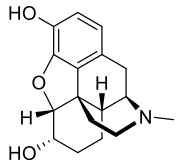


Figure 1.1: Chemical structure of morphine, currently marketed under many trade names.

The late 19th century saw sufficient advances in chemical science to isolate compounds and determine chemical structures. The Merck Group (est. 1668) was one of the first companies to move into large scale drug production. Until the mid-twentieth century drugs were discovered primarily by testing analogues of known drugs and their metabolites on animal models. Chemical and molecular biology revolutionised the field of drug discovery in the late 1980's via the rapid improvements in understanding the molecular basis of disease. The introduction of computational modelling and rational chemical methods improved the drug design process. New technologies, such as high throughput screening (HTS) and combinatorial chemistry enhanced the discovery process by being able to screen millions of compounds, and the number of emerging technologies continues to grow.[2] The modern drug discovery process involves the identification of screening hits and subsequent optimisation by medicinal chemistry to increase the pharmacological profile of the final drug compound.

1.1.2 Drug-like Compounds and Properties

The drug discovery process is plagued with high attrition rates; in a typical drug discovery process > 90% of drug candidates fail to reach market.[3] The main reason for this is that many new drugs have no efficacious advantage over existing ones. Drug molecules have to have sufficient activity and efficacy at the site of action to moderate the target. There are various

metrics used to optimise the compounds during the drug discovery process, the most important of which are described in table 1.1.

Metric	Description
Affinity	The strength of the interaction between the ligand and protein. The dissociation constant, K_d, is the most common measure of the binding strength. A low K_d = high affinity.
Efficacy	A measure of the maximum response achieved with an applied dose of drug compound.
Selectivity	The ability of a drug to preferentially target a particular a site of action, relative to other sites.
Pharmacokinetics	The effect the body has on a drug compound, also referred to as ADME (Absorption, Distribution, Metabolism and Excretion).
Toxicity	This is a broad description, however, mainly refers to how harmful the drug is at therapeutic doses.

Table 1.1: The various metrics used to characterise a drug compound during development.

There are two main, distinct types of drug molecules used in development of clinical candidates: peptides and small molecules. The following sections introduces both peptides and small molecules and focusses on the main differences and the medicinal potential of each.

1.1.3 Small Molecules

It is without question the drug market is dominated by small molecule (SM) compounds. In 2014, five of the top ten selling drugs were small molecules[4]: aripiprazole (Abilify, $7.2 billion), esomeprazole (Nexium, $6.3 billion), rosuvastatin (Crestor, $5.6 billion), fluticasone propionate (Advair Diskus, $5.0 billion), and sofosbuvir (Sovaldi, $4.4 billion). The SM class of drug compounds represents an almost infinitely diverse set of molecules that are waiting to be discovered.

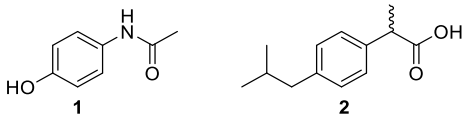


Figure 1.2: Chemical structures of **1**) paracetamol and **2**) ibuprofen.

Two of the most recognised and simplest examples of SM drugs are paracetamol and ibuprofen (Figure 1.2). These are the most widely available over the counter analgesics in the

world today, despite being available since the 1950's.[5] However, SM drugs continue to lead the way in number of new compounds brought to market, a leading example of a modern SM drug is Vertex Pharmaceuticals' Ivacaftor (Figure 1.3). Approved in 2012, it is the first drug to treat the underlying cause of patients with cystic fibrosis.[6]

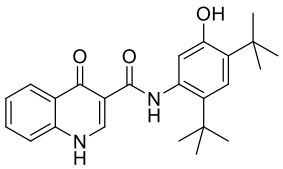


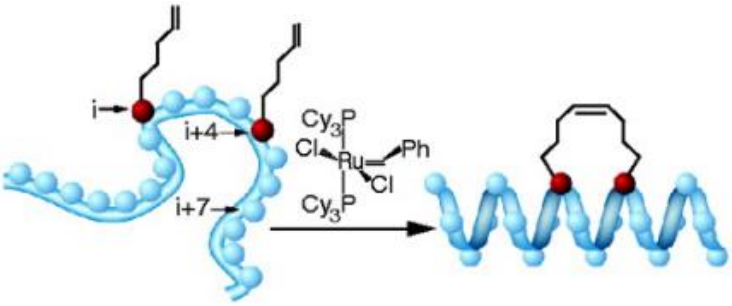
Figure 1.3: Chemical structure of Ivacaftor. Used to treat patients with cystic fibrosis caused by certain mutations in the CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) gene.

Classic drug development of SMs involves synthesis by well-defined chemical reactions and they can be easily characterised by standard analytical methods. Also, ADME testing of SMs is generally much easier (than peptides) due to predictable physiochemical properties and a reduced structural complexity.[7] They are also preferred, over peptides, primarily due to ease of production, simplicity of administration (oral availability) and their superior pharmacological properties to peptides. However, the high financial expenses associated with SM drug development has reduced the number of approved drugs reaching the market. Currently, the pharmaceutical industry is undergoing a paradigm shift towards peptides as therapeutic agents coinciding with the advances in molecular and synthetic biology.[8]

1.1.4 Peptide Therapeutics

Recently, there has been a substantial increase in the use of peptides as therapeutic agents.[9] In general, peptides are typically short-chains of amino acids that bind to cell surface receptors, such as GPCRs or ion-channels, where they mediate intracellular signalling pathways. A host of functions have been identified for naturally occurring peptides including: hormones, growth factors, neurotransmitters, and ion channel ligands.[8] Their high target specificity, efficacy and safety generally makes them a more attractive choice of treatment than small molecule drugs.[10] However, naturally occurring peptides are generally unsuitable due to suboptimal chemical and physical stability, and a poor bioavailability.[11] Therefore, rational techniques to alleviate these issues is a large area of research. The motivation for pursuing peptides as therapeutics is the ease of synthesis and scope for novel modifications. A traditional structure-based design strategy to establish structure-activity relations (SAR) of

natural peptides is called 'alanine scanning'. This is an approach used to identify the amino acids in a known peptide sequence that are essential for the peptide's stability and function. Sequential replacement with an alanine residue in the peptide sequence indicates the contribution of each residue to the binding interaction.[12] In this way, peptides can be modified with, for example, non-natural amino acids to advantageously alter the chemical and physical properties, such as stability and biological activity, these are called peptidomimetics. In 2004, Walensky *et al* reported the synthesis of a 'stabilised alpha-helix of BCL-2 domains' (SAHBs) which mimic the BH3 segment in BCL-2 regulated apoptosis.[13] Covalent helix stabilisation was achieved by replacement of two amino acids in the backbone of a known α-helix 23*mer* with olefinic containing amino acids and subsequent cross-linking by ruthenium-catalysed olefin metathesis (Figure 1.4). This led to an enhanced affinity, helicity, protease resistance and cell permeability of the peptidomimetic demonstrating the benefits of non-natural amino acids.



Compound	**Sequence**
Native BH3 peptide	EDIIRNIARHLAQVGDSN$_L$DRSIW
SAHB$_A$	EDIIRNIARHLA*VGD*N$_L$DRSIW

* = olefinic containing amino acid, N$_L$ = norleucine

Figure 1.4: Top: Covalent helix stabilisation of an α-helix by inclusion of non-natural amino acids and subsequent crosslinking reaction. Bottom: table shows the native sequence and synthetic sequence of the final peptidomimetic. Adapted from Walensky, 2004.[13]

There are many examples of including non-natural amino acids to modify peptide properties, a greater description of methods and screening technologies related to peptides is given in chapter 2.

1.1.5 Drug Discovery Pipeline

A modern view of the drug discovery process can be illustrated by a drug pipeline (Figure 1.5). It consists of five major stages: target identification, hit generation, hit-to-lead, lead optimisation and preclinical development. This is a broad overview and ultimately depends on the nature of the project, e.g. compounds based on known drugs already have a validated target.

Figure 1.5: Drug Discovery Pipeline. The first stage of a drug discovery campaign is to identify a target. This is followed by hit identification and optimisation of this hit into a potent lead compound. At this stage the compound enters preclinical development followed by clinical trials.

There are a plethora of challenges facing the pharmaceutical industry in development of new therapeutics, ranging from antibacterial resistance to translating *in vitro* efficacy to animal models.[14] The major challenge facing modern drug discovery is an insufficient knowledge of the molecular mechanisms of disease, therefore discovery is hindered at the first hurdle. This thesis is aimed at the very first stages of the drug discovery process by development of new approaches to generate a hit compound for difficult, orphan targets. Orphan protein targets include those that are generally under-explored in academia and pharma industry and have no known inhibitors and are therefore considered difficult or undruggable. The PPI protein targets presented in thesis are detailed in chapters 2 and 4.The following sections describe the first two stages of the drug discovery pipeline.

1.2 Target Selection and Validation

The first step towards developing a drug is to identify the druggable genes at the root of disease. Most drugs currently on the market have a known target for interaction, and there are various methods of identifying (determining a drug target) and validating (demonstrating the effect of the interaction) the drug target. These two steps are closely associated and are generally considered together. Most drug targets are proteins, however, DNA and RNA can also be the focus of drug intervention. The advances in structural genomics and analysis of sequence-structure homology has generated a huge influx of new protein targets for

validation. However, the general principle of validation has to be considered throughout the entire drug discovery process and relies on: cell-based validation, validation in animal studies then proof-of-concept in humans (clinical trials). The methods towards target validation can be broadly split into two categories: biological and chemical, as illustrated in table 1.2. Biological validation generally adopts a functional *in vivo* assay (phenotypic observations) to determine target validity, whereas as chemical validation focusses on mechanistic studies of binding interactions *in vitro*.

Biological (*in vivo*)	**Chemical (*in vitro*)**
• Gain of function: gene over-expression in cells or whole animals • Loss of function: gene knockout tests in model organism	• Protein inhibition by small molecules, peptides, antibodies and aptamers in chemical assays (affinity determination).

Table 1.2: This table gives an overview of the biological and chemical methods of target validation.

1.2.1 Target Focus in Drug Discovery

There are approximately 30,000 genes in the human genome, leading to the concept of the druggable genome.[15] Hopkins and Groom (2002) carried out an extensive analysis of the number of molecular targets amenable to drug intervention.[15] It was found that there were 130 distinct protein families representative of known drug targets, however, nearly half of all targets fell into just six families: G-protein-coupled receptors (GPCRs), serine/threonine and tyrosine protein kinases, zinc metallopeptidases, serine proteases, nuclear hormone receptors and phosphodiesterases. A more recent class of therapeutic targets, which were deemed intractable twenty years ago, are protein-protein interactions, and these are discussed in the next section.

1.2.2 Protein-Protein Interactions

Many of the fundamental processes in cells are controlled by proteins interacting with one another. These are called protein-protein interactions (PPIs) and are critical in the regulation of biological systems and controlling signal transduction pathways. The dysregulation of this vast interactome in the development of disease is widely recognised. However, inhibitors of these PPIs are relatively scarce, as these are generally considered difficult, or undruggable,

targets. Peptide inhibitors are often used to target PPIs due to their native affinity for proteins, however, as described previously, they are hindered by their inferior stability and bioavailability to small molecules. Despite this, recent work has emerged demonstrating that certain PPIs are amenable to small molecule inhibition by a similar mode of action to peptides, by binding into pockets on the surface of proteins and disrupting the interaction. Further mechanisms of small molecule inhibition of PPIs include: allosteric mechanisms or binding at a catalytic site. High resolution structures, gleaned from X-ray crystallography or NMR, has shown that PPI interfaces are generally flat and large (1,000 – 2,000 $Å^2$)[16], however, most inhibitors target PPIs where the essential binding interactions are concentrated in 'hot-spots'. These hot-spots are the interactions responsible for driving the affinity of two proteins and are specific residues distributed across the surface of the proteins. PPIs can be generally classified by whether the interface consists of a primary protein sequence, a single region of secondary structure or multiple sequences containing tertiary structures, this is illustrated in figure 1.6.

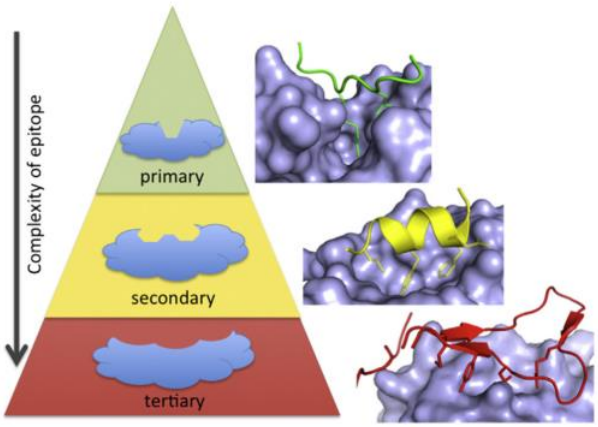


Figure 1.6: A diagram to show the decreasing druggability of a PPI according to the complexity of the epitope. Image taken directly from Arkin *et al*.[17]

As described earlier (section 1.1.4), the alanine scanning technique was one of the leading methods that inspired confidence that PPIs could be modulated by small molecules. In the last decade, more than 40 PPIs have been targeted and there are several inhibitors currently in clinical trials, this makes it the fastest growing target class for drug inhibition.[18]

1.2.3 Inhibitors of PPIs

One of the first examples of a PPI inhibitor approved for clinical use is Tirofiban, an antiplatelet drug used to reduce the rate of thrombotic cardiovascular events. It is a small molecule designed to mimic the linear peptide RGD, which is the epitope of fibrinogen that

interacts with glycoprotein IIb/IIIa (figure 1.7). This is a good example of the design of a small molecule from a known peptide inhibitor using prior knowledge of the essential binding interactions.

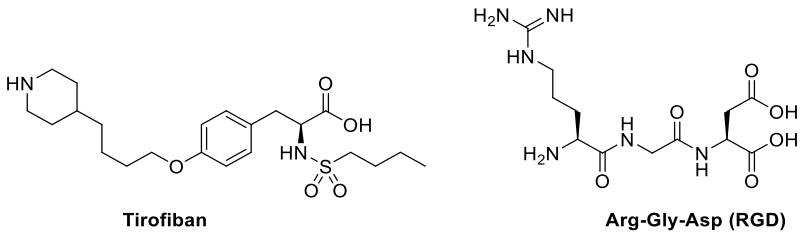


Figure 1.7: Chemical structure of Tirofiban and the mimicking peptide (RGD).

Arguably the largest area of drug research is development for cancer therapies. Inhibitor of apoptosis proteins (IAPs), including XIAP (X-linked Inhibitor of Apoptosis), are regulators of apoptotic cell death.

These IAPs have interaction domains called BIR domains that directly bind to an N-terminal tetrapeptide sequence on caspases and thereby prevent caspase-promoted cell death. An example of an endogenous inhibitor of IAPs is Smac (second mitochondrial activator of caspases), which competes with caspases to stimulate apoptosis. Therefore, small molecule inhibitors are designed to mimic the tetrapeptide binding motif on Smac and are called Smac mimetics. Several Smac mimetics have been developed that show anti-tumour activity, these are summarised in table 1.3.

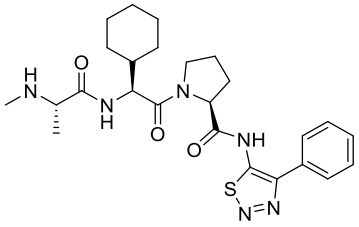
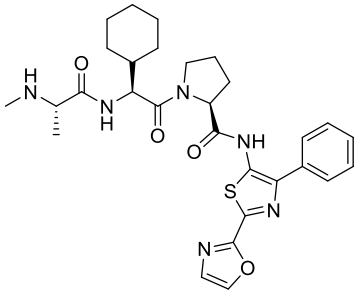
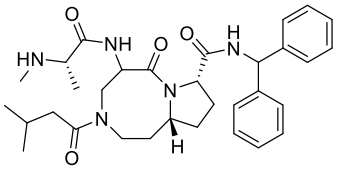
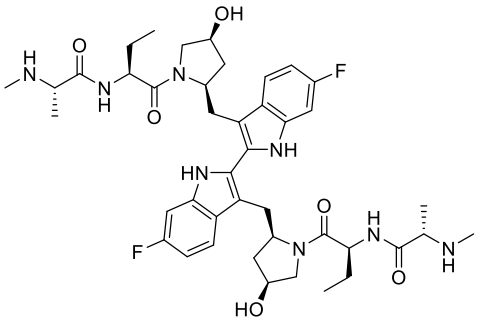
ID	Company	Structure
GDC-0152	Genentech	
GDC-0917	Genentech	
AT-406	Ascenta	
TL-32711 (Birinapant)	TetraLogic	

Table 1.3: A table showing various SMAC mimetics (peptide-like inhibitors of primary epitope PPIs) currently in clinical trials. Adapted from Arkin *et al*.[17]

1.3 Hit Discovery

After a target has been selected and shown to have a relative capacity for intervention, the next stage is to find ligands for the target. This is generally done by application of screening campaigns in industry and there are two, distinct alternative approaches to hit discovery: high throughput screening (HTS) and fragment-based drug discovery (FBDD). These are illustrated in figure 1.8 and discussed in the following sections.

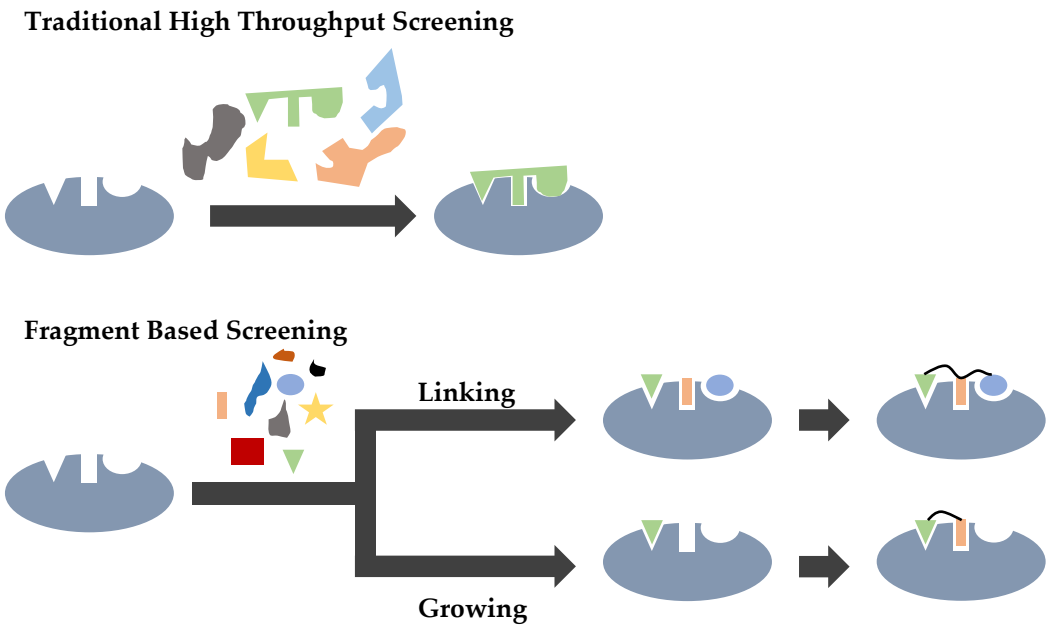


Figure 1.8: A schematic diagram of the two approaches to hit discovery: HTS and FBDD.

1.3.1 High Throughput Screening - HTS

Originally, the pharmaceutical industry used high throughput screening to discover new lead molecules against a chosen target. This consisted of screening many thousands (50,000 – 1,000,000) of large, complex, drug-like molecules for their ability to modify the biological activity of the chosen target.[19] Companies have invested in complex hardware and robotic systems to test large numbers of compounds quickly and efficiently and provide a competitive advantage.[20] The HTS method is frequently employed with analytical techniques such as X-ray crystallography, NMR or coupled methods, e.g. LC-MS/MS. Also, biophysical techniques such as fluorescence spectroscopy and isothermal titration calorimetry have also been used to screen compound libraries. However, an important caveat of HTS is that it requires a robust assay specific to the target to minimise the effect of false positives.[21] This is time-consuming to develop and requires detailed knowledge of the biological function of the target. This also requires the use of orthogonal techniques to validate the hit compound.

Over the past two decades the discovery and development of new drug compounds has relied heavily on these HTS techniques. The pharmaceutical industry has invested heavily in HTS projects that have yielded fewer and fewer drug leads and clinical candidates particularly against difficult targets. Multimillion member compound libraries deliver poor hit rates or unsuitable compounds, therefore, there is a clear need to deliver alternative approaches to

address these limitations. One approach that is becoming increasingly popular in industry and academia is fragment-based drug discovery (FBDD).

1.3.2 Fragment Based Drug Discovery - FBDD

"Great things are done by a series of small things brought together" – Vincent van Gogh.

Fragment-based drug discovery (FBDD) is now a well-established approach in academia and the pharmaceutical industry for the discovery of ligands that bind to proteins and/or nucleic acid targets.[22] This method is used extensively in the search for new drug compounds, with many already in clinical trials[23], and with the first fragment-derived compound treating patients[24].

The concept of the FBDD approach is that large drug-like molecules are built up of two, or more, smaller, lower complexity components, i.e. fragments. These small fragments can sample a larger volume of the available chemical space. Chemical space is a rather abstract concept that describes the theoretical number of molecules that can be synthesised, numbers range from 10^{60} - 10^{100} molecules.[25] Screening for the smaller fragments using low molecular mass, fewer compound libraries (in the range of hundreds to a few thousand) will yield a core structural starting point for hit-to-lead optimisation. Screening fewer compounds is also faster, less expensive and hit fragments have the potential to keep complexity and molecular mass low of lead compounds. However, detection and characterisation of such weak affinity compounds is problematic using many traditional biochemical HTS assays. Therefore, highly sensitive biophysical screening methods, such as NMR or X-ray crystallography are required to detect binding (see section 1.3.3). A crucial aspect to FBDD is in the design of the library and choosing suitable compounds for screening, this is discussed in further detail in chapter 3.

There are two options for hit development after fragment screening, these are illustrated in figure 1.8. Fragment linking is an attractive option to increase the potency of two hit fragments to a particular target, this was demonstrated in the 'original' SAR by NMR approach implemented by Fesik *et al*.[26] Fragment growing requires extensive medicinal chemistry to optimise hit compounds in the hit-to-lead development stage. This is costly and time-consuming however, generally, simpler chemical syntheses are required over complex hits obtained in HTS. The FBDD approach is also impacting significantly on hit discovery for

Status	Drug	Company	Target
Approved	Vemurafenib	Plexxikon	B-Raf(V600E)
	Venetoclax	AbbVie/Genentech	Selective Bcl-2
Phase III	PLX3397	Plexxikon	FMS, KIT, and FLT-3-ITD
	Verubecestat	Merck	BACE1
	AZD3293	AstraZeneca/Astex/Lilly	BACE1
Phase II	AT7519	Astex	CDK1,2,4,5,9
	AT9283	Astex	Aurora, JAK2
	AZD5363	AstraZeneca/Astex/CR-UK	AKT
	Erdafitinib	J&J/Astex	FGFR1-4
	Indeglitazar	Plexxikon	pan-PPAR agonist
	LY2886721	Lilly	BACE1
	LY517717	Lilly/Protherics	FXa
	Navitoclax (ABT-263)	Abbott	Bcl-2/Bcl-xL
	NVP-AUY922	Vernalis/Novartis	HSP90
	Onalespib	Astex	HSP90
Phase I	ABL001	Novartis	BCR-ABL
	ABT-518	Abbott	MMP-2 & 9
	ABT-737	Abbott	Bcl-2/Bcl-xL
	ASTX660	Astex	XIAP/cIAP1
	AT13148	Astex	AKT, p70S6K, ROCK
	AZD3839	AstraZeneca	BACE1
	AZD5099	AstraZeneca	Bacterial topoisomerase II
	BCL201	Vernalis/Servier/Roche	BCL-2
	DG-051	deCODE	LTA4H
	IC-776	Lilly/ICOS	LFA-1
	LP-261	Locus	Tubulin
	LY2811376	Lilly	BACE1
	PF06650833	Pfizer	IRAK4
	PLX5568	Plexxikon	Kinase
	SGX-393	SGX	BCR-ABL
	SGX-523	SGX	Met
	SNS-314	Sunesis	Aurora

Table 1.4. This table compiles the most recent (2016) fragment-derived drugs currently in clinical development. Table adapted from Erlanson, 2016.[28]

challenging targets, such as protein-protein interactions (see section 1.3.2), where many HTS campaigns are unsuccessful.

Currently, FBDD is practiced widely in academia and industry and a recent review, 2016, of fragment-based drug design includes an up-to-date list of fragment-derived drugs that are in clinical development, these are shown in table 1.4. Arguably one of the most eagerly anticipated drugs included on this list is the BACE-1 (Beta-secretase 1) inhibitor Verubecestat from Merck. The so-called 'Amyloid hypothesis' is a leading philosophy of neurodegeneration in Alzheimer's disease. Its primary directive involves the accumulation and deposition of amyloid β-peptide (Aβ) as plaques in brain tissue.[27] Key to the formation of these short, aggregating peptides is cleavage of the amyloid precursor protein (APP) by BACE-1, therefore, in theory, an inhibitor of this membrane protein would prevent build-up of Aβ and may slow or prevent Alzheimer's disease. This shows that FBDD is a significant technique for the identification of new hit compounds for difficult, orphan targets and diseases. A summary of the advantages and disadvantages of HTS and FBDD is given in table 1.5.

	Fragment Based Screening	**High Throughput Screening**
Advantages	• High hit rates (3 – 10 %) • Fast • Cost effective • Hits are easier to optimise	• Large numbers of compounds can be screened quickly • High affinity binders • Minimisation and low complexity of assay • Reduction in costs
Disadvantages	• Low affinity binders • Sensitive detection methods required • Extensive medicinal chemistry	• Low hit rates (< 1%) • Complex medicinal chemistry • High attrition rates in lead development

Table 1.5: A comparison of the advantages and disadvantages of HTS vs FBDD.

1.3.3 Methods in Fragment Based Drug Discovery

A fundamental aspect to FBDD is the chosen assay to screen the compound library. Fragments are inherently weak affinity binders, therefore highly sensitive methods are required. Among the classical, biophysical assays for fragment screening are NMR spectroscopy, X-ray crystallography and SPR. In a recent poll, 2016, on the popular drug discovery blog '*Practical*

Fragments' the most popular technique used by respondents for FBDD was ligand-detected NMR (Figure 1.3.1), followed by SPR and X-ray crystallography.[29]

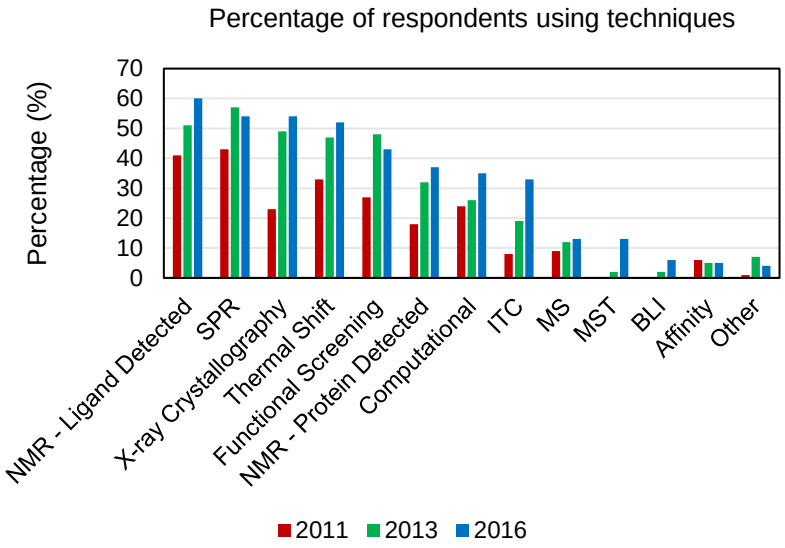


Figure 1.9: A chart showing the most popular techniques used in FBDD. Adapted from Erlanson [23].

Method	Sensitivity Range	Description
Ligand-observed NMR (STD, WaterLOGSY, CPMG)	10 mM – 100 nM	Can detect very weak binders however requires large amounts of protein. Advantageous to confirm ligands bind in solution
Protein-observed NMR (HSQC)	5 mM – 100 nM	Requires isotopic labelling of the protein (size limit < 40 kDa). Affinity determination and binding site mapping can be determined.
X-ray crystallography	All affinites	Crystal structure provides detailed description of binding site and essential interactions. Co-crystallisation can required different crystal conditions for each ligand, not very amenable with organic solvents.
Surface plasmon resonance	500 μM lower limit	Allows kinetics of binding to be measured and accurate K_d determination. Difficult to immobilise protein.

Table 1.6: Summary of classical FBDD screening methods. Adapted from Hubbard.[24]

Arguably, the origin of FBDD began with the work of Fesik and co-workers in their paper entitled 'SAR by NMR', and the use of NMR in drug discovery projects has greatly increased. Table 1.6 summarises the classical methods utilised in FBDD. NMR is advantageous as it requires almost no assay development, all that is needed is a soluble target and i.e. there is no requirement for crystallisation, immobilisation or labelling of the target. The practical applications of NMR in FBDD are discussed in more detail in section 1.4.

1.3.4 Fluorine in Medicinal Chemistry

Fluorine chemistry is a large area of research, particularly in the pharmaceutical and biomedical industries. Advances in synthetic organic chemistry has made possible the production of a wide range of fluorine-containing analogues of many different biological compounds, e.g. nucleosides, amino acids, lipids and sugars. The role of fluorine in drug design is increasing and is now widely recognised for its unique properties. The inclusion of a fluorine atom in a drug compound can have advantageous effects on a molecule's physiochemical properties, including:[30–32]

- pK_a: a fluorine atom is generally introduced to increase the acidity of acids and decrease the basicity of bases.
- Lipophilicity: is also usually increased if a fluorine atom is placed in the vicinity of basic nitrogens or introduced to aromatic rings. Lipophilicity is particularly increased by the introduction of trifluoromethyl or trifluoromethoxy (CF_3 or OCF_3 respectively) containing groups, whereas a single F atom (CF) may alter it in either direction.
- Membrane permeability: can be increased by introduction of an F atom at a specific site of a molecule due to intramolecular hydrogen bonding
- Metabolic stability: a fluorine atom is frequently used to replace a H atom in a lead compound where metabolism is likely to occur. The higher dissociation energy of a C–F bond, compared to a C–H bond
- Binding affinity: a single fluorine atom in a molecule can also have a large effect on the binding affinity of the compound to the target. This is due to the different van der Waals radius' and stereoelectronic effects of fluorine and hydrogen and their specific interaction with the target, for example, hydrogen bonding interactions and lipophilic contacts.

Historically, Fludrocortisone was the first example of a fluorinated drug compound to be developed (Figure 1.10) and many of the currently available market leading pharmaceuticals now contain fluorine atoms,[33] as shown in table 1.7.

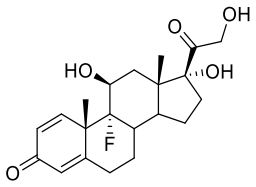


Figure 1.10: Chemical structure of Fludrocortisone, the first fluorinated pharmaceutical to be developed.

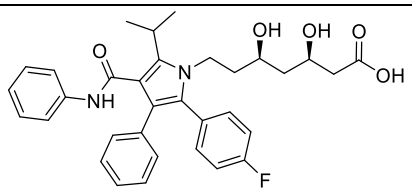
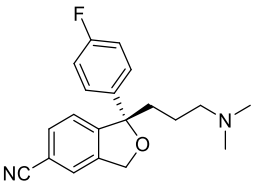
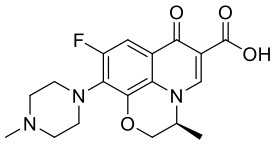
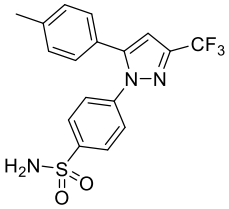
Drug	Structure	Treatment
Lipitor (Atorvastatin)		Inhibits cholesterol biosynthesis
Lexapro		Depression/anxiety
Levaquin		Antibiotic
Celebrex		Anti-inflammatory

Table 1.7: Structures of some fluorine-containing drugs and their uses.

This thesis focusses on the use of fluorine containing compounds for screening by fluorine NMR methods. A comprehensive review and approach to fluorinated peptides and fluorine fragment library design is described in chapters 2 and 3, respectively.

1.4 Applications of NMR Spectroscopy in Drug Discovery

Nuclear magnetic resonance (NMR) has found many applications in FBDD. NMR can be used for screening large libraries and is an important alternative to X-ray crystallography in determining high resolution structures. Over the years, several NMR methods have been developed towards fragment-based screening and these techniques can be categorised into two groups: ligand-observed NMR and protein-observed (target-based) NMR, these are discussed in the following section.

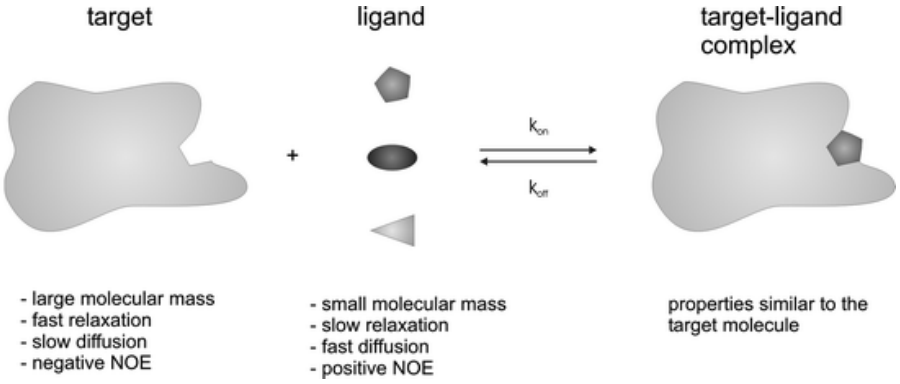


Figure 1.11: NMR properties of proteins and small molecules. Fragments adopt the properties of the target protein upon binding due to the increase in the effective molecular mass. From Klages *et al*.[34]

1.4.1 Protein-observed Screening Methods

Almost exclusively, protein-observed NMR methods measure chemical shift perturbations (CSPs) to detect ligand binding. The chemical shifts of certain protein residues in the binding site will change upon the addition of ligand due to changes in the local environment. The most common protein-observed NMR technique used is the 2D heteronuclear single quantum coherence (HSQC) experiment. The isotopic labelling with ^{13}C or ^{15}N allows for the identification of residues involved in the binding interaction which can then be used for affinity determination or binding site mapping. Again, the original 'SAR by NMR' approach adopted by Fesik *et al* observed CSPs of ^{15}N-HSQC NMR spectra of However, isotopic labelling is relatively expensive and protein expression typically produces poor yields. Also, 2D spectra take much longer to acquire than 1D and a relatively high concentration of protein is required to obtain a practical signal to noise (S/N) ratio. A further limitation of protein-

observed NMR is the effective size of the protein. For large macromolecules, the number of signals increases and they become broader due to slower molecular tumbling. This results in spectral overlap which can complicate spectral analysis. A ^{1}H-^{15}N correlation is a molecular fingerprint of the protein and is a good measure of whether the protein is folded or aggregated.

1.4.2 Ligand Observed Screening Methods

Ligand-observed methods are based on the fact that small molecules adopt the properties of the protein upon binding (Figure 1.11). For example, proteins exhibit fast relaxation rates, positive NOE (Nuclear Overhauser Effect) cross correlation and slow diffusion coefficients, whereas small molecules have slow relaxation, negative NOEs and large diffusion coefficients. Ligand-observed methods generally measure a 1D spectrum and are therefore faster than protein-observed methods. Also, it is possible to screen large mixtures of ligands as the ligand is directly detected, therefore no deconvolution of the hit compound is required. Measuring the spectra in compound pools is faster and often requires lower amounts of protein. Another advantage of ligand-observed methods is there is no need to isotopically label the protein and often higher molecular weight proteins can be screened. The most common NMR method for ligand-observed screening is saturation transfer difference (STD) NMR. Mayer (1999) was the first to apply the STD technique to screening of protein-ligand interactions. The general principle involves subtracting two spectra, one recorded in which the protein was selectively saturated (on-resonance) from one recorded without protein saturation (off-resonance) (Figure 1.12). The on-resonance spectrum is obtained by irradiating at a region that contains only resonances of the protein target (typically -1 – 0 ppm; $–CH_3$ methyl protons) which gives signal intensities I_{SAT}. The off-resonance spectrum is obtained by irradiating at a frequency where no ^{1}H signals occur, with signal intensities I_0. During acquisition of the on-resonance spectrum, saturation is transferred through the protein via spin-diffusion and any bound ligands will also receive saturation. The two spectra are subtracted from one another and any bound ligands will be visible due to the difference in intensities, non-binding ligands have the same intensity in each spectrum and are therefore cancelled out. The advantages of STD NMR include: use of an excess of ligand which allows for low consumption of protein and no upper size limit of the protein.

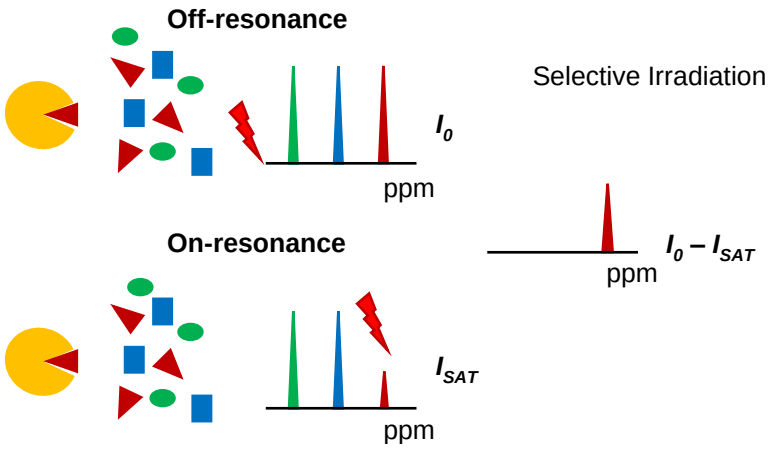


Figure 1.12: Scheme showing the two required spectra for the Saturation Transfer Difference-NMR experiment. The selective saturation of protein resonances allows intermolecular transfer of magnetisation from the protein to any bound ligands. Adapted from Viegas *et al*.[35]

However, there is a lower size limit (about 10 kDa) as small proteins are not as effectively saturated by spin diffusion. A recent example (2016) of fragment-based screening by STD NMR is by Mesleh *et al* who screened a library 5643 fragments against *S. aureus* DNA Gyrase B which yielded 304 hits.[36] Fragment hits were initially classified based on the STD signal intensity and then further reduced based on fragment novelty and chemical tractability, yielding a total of 46 fragments for follow up.

Another popular method for NMR screening, similar to STD, is water-ligand observation by gradient spectroscopy (WaterLOGSY). The difference between STD and WaterLOGSY is that magnetisation is transferred from bulk water molecules to the protein via two pathways: 1) transfer from water molecules in the binding pockets of the protein and 2) hydrogen exchange of chemically labile hydrogens on the protein. Water molecules are almost always found linking ligands to the protein via a network of intermolecular hydrogen bonds.[37] Therefore, magnetisation is transferred to any bound ligands via the water molecules in the binding site. Understandably, magnetisation will also be transferred to non-bound ligands in solution however, bound and non-bound ligands can be distinguished due to the difference in sign of the NOE. The large, protein-ligand complex tumbles slowly and has a negative NOE, however, small molecules rotate fast in solution and have a positive NOE. Therefore, protein bound ligands show a positive signal and non-binding ligands show a negative signal in the difference spectrum. WaterLOGSY is a sensitive method and can be used for mixtures of compounds and is especially useful for RNA binding proteins as these are typically strongly hydrated.

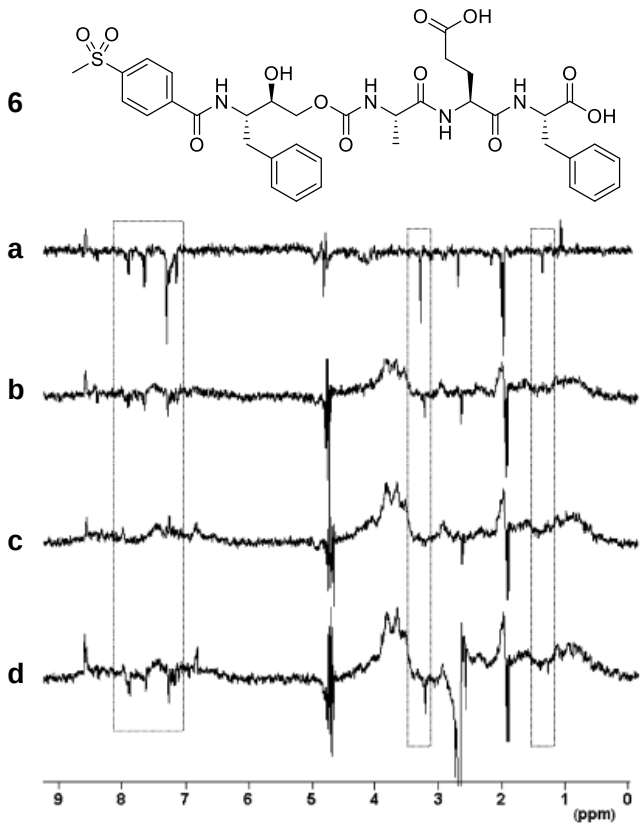


Figure 1.13: A series of NMR WaterLOGSY experiments showing the binding of compound 6 (IC_{50} 30 μM) to BACE-1. **A**) No protein (negative peaks for unbound compound); **B**) Spectrum after addition of 1.8 μM BACE-1 (peaks are less negative). **C**) With 3.2 μM protein (compound signals are increasingly positive). **D**) Addition of high affinity ligand to displace compound 6 (NMR signals are negative again). Adapted from Geschwindner *et al*.[38]

1.4.3 Fluorine NMR Spectroscopy for Screening

There are many advantages of using the fluorine nucleus as an NMR probe for investigating biological systems.[39] The ^{19}F nucleus has a nuclear spin of ½, a high gyromagnetic ratio, 100% natural abundance and a sensitivity that is 83% that of the proton and is therefore ideal for studying protein-ligand interactions by NMR spectroscopy.[39] It is widely used to study protein-ligand interactions either through ligand or protein observation, or indirectly through a reporter molecule. In the context of drug discovery and screening of large compound libraries, ligand-observed NMR methods are usually preferred, although protein-observed methods are not uncommon.[40,41] In contrast to ^{1}H observed ligand methods, ^{19}F-based experiments routinely measure chemical shift perturbations (CSPs). The chemical shift range is very large, compared to the proton, and is very sensitive to the local electronic environment and to the changes during a ligand binding event. Furthermore, due to the high chemical shift anisotropy (CSA) of the fluorine nucleus, this makes it sensitive to strong line broadening

effects at slow molecular tumbling rates, for example during a ligand binding to a large protein.[40] Thus, the fluorine nucleus is ideal for use as a sensor to detect protein-ligand interactions.

^{19}F NMR is a relatively new approach to fragment screening and there are various methods to detect fragment binding currently employed.[42,43] Fluorine NMR spectra are much simpler than ^{1}H spectra since there are typically fewer signals and no signals from the protein. This allows for screening of large mixtures of compounds, compared to proton observed methods, which greatly reduces the measuring time.[44] Typically, a few thousand molecules can be screened per day and only a small amount of unlabelled protein is required.[45] The main concern regarding fluorine fragment-based screening is the available chemical space coverage of fluorine containing compounds, this is discussed in more detail in Chapter 3. An excellent review in Drug Discovery Today from Anna Vulpetti and Claudio Dalvit, describes the ubiquitous use of ^{19}F methods in drug discovery.[46] They propose a complete workflow of the various ways that fluorine can impact the hit discovery and validation stages. The main extension and utility of fluorine NMR-based methods described is in competition-based assays using ^{19}F-containing reporter molecules.[47] For example, screening of a rudimentary ^{19}F-based fragment library can lead to weak affinity ligands. These are then used as 'spy' molecules in a competition-based NMR experiment termed FAXS (Fluorine Anisotropy and Exchange for Screening).[40,48] Further methods of ^{19}F NMR-based screening and the practical aspects are discussed in more detail in chapter 4.

1.5 Concluding Remarks

In summary, drug discovery is a vast area of research and the identification of potential drug candidates can be achieved by a plethora of techniques. A crucial aspect to the success of a drug discovery campaign is in design of the library, whether it is peptides or small molecules, a review of library design strategies is given in chapter 3.

NMR screening has become a powerful tool for the discovery and validation of drug candidates in the pharmaceutical industry. Advances in NMR technology has made screening applications involving NMR indispensable to modern drug discovery. NMR techniques for screening do not require prior knowledge of the protein's function and can therefore be applied to targets for which no biochemical assay is available. This is ideal for efficient drug discovery against difficult target such as PPIs as described in this thesis. Also, NMR

techniques are solution-based methods, therefore are not compromised by crystallisation conditions or target immobilisation. The K_D detection range is also very large and is in the ideal range for fragment-based screening where hits are likely to be around the μM to mM range. Also, NMR allows QC of fragments during setup and identification of poorly behaving compounds (i.e. aggregators) is straightforward. As described previously, the inclusion of a fluorine atom serves multiple purposes in drug discovery projects. The use of fluorine as a detection tool for ligand-observed NMR screening is rapidly gaining momentum for fast, efficient drug discovery. The aims of this thesis are to discover new, novel ligands for difficult, underexplored protein targets by ^{19}F NMR-based screening and strategic design of peptide and fragment libraries. The design, curation and evaluation of these methods are explored

Chapter 2

Peptides and Peptidomimetics

This chapter contains the results of work performed on OBOC synthesis of peptide libraries and SPPS of fluorinated peptides for screening purposes. A short introduction covers the background to the experimental approach and also information on the protein targets for screening. The results and discussion section is split into two parts: 1) peptide tools for ATAD2 targeting and 2) a novel 'on-protein peptide growing' approach. Each part closes with a conclusions section.

2.1. Introduction

2.1.1. Towards the use of Peptides as Screening Agents

Over the last two decades, high throughput screening of large libraries of compounds against the most validated targets has been the method of choice for efficient drug discovery. One important aspect for success in the early drug discovery process relies on screening the right classes of chemical compounds for a specific target. Highly diverse small molecular libraries are extremely costly and logistically difficult to handle.[49] While libraries of up to 5 million compounds (e.g. Pfizer) were successfully screened against enzymes and GPCRs, the success of finding chemical starting points for PPIs was very limited.[50] Besides of the fragment screening approach covered in chapter 3 and 5 of this thesis, the second most successful and still promising chemical reagent class are peptides and peptidomimetics. To discover a novel lead compound, whether it is a small molecule or peptide, has led to considerable effort in developing alternative fast and flexible screening methods. Two different approaches for finding hit compounds for difficult targets are covered in the sections below; one of them, to our knowledge has not been applied before.

Peptides are a central focus in biological, medicinal and pharmaceutical research; therefore methods of synthesis are a major focus in organic chemistry. Solid-phase synthesis of peptide chains was first reported by Merrifield in 1963, and for development of this technique he received the Nobel Prize in 1984.[51] Merrifield's work revolutionised the field of peptide chemistry, facilitating the production of high molecular weight polypeptides in high yield and purity thereby eliminating the need for purification. The initial Merryfield technique of peptide synthesis was developed further towards generating small peptide libraries of known sequences.[52] Peptide libraries can be used as targeted and focused, or as a "random" screening library. One-bead one-compound (OBOC) library synthesis and screening by various methods in solution and on the solid surface, allows efficient ligand discovery for many targets and many applications, most importantly in drug discovery and molecular imaging.[53]

2.1.2. One-Bead One-Compound (OBOC) Libraries

In 1991, Lam *et al* developed a peptide screening approach named 'one-bead one-compound' (OBOC) synthesis.[54] In the OBOC approach, compounds are prepared by solid-phase synthesis in a "split and mix" technique, resulting in the display of many copies of a

compound on one single micro bead (figure 2.1). Modern solid-phase combinatorial techniques have achieved efficient synthesis of compounds each containing ~100 pmol of a single compound per bead.[55] This method can produce tens of thousands to millions of compounds within a single library at relatively low cost and short timescales.[53]

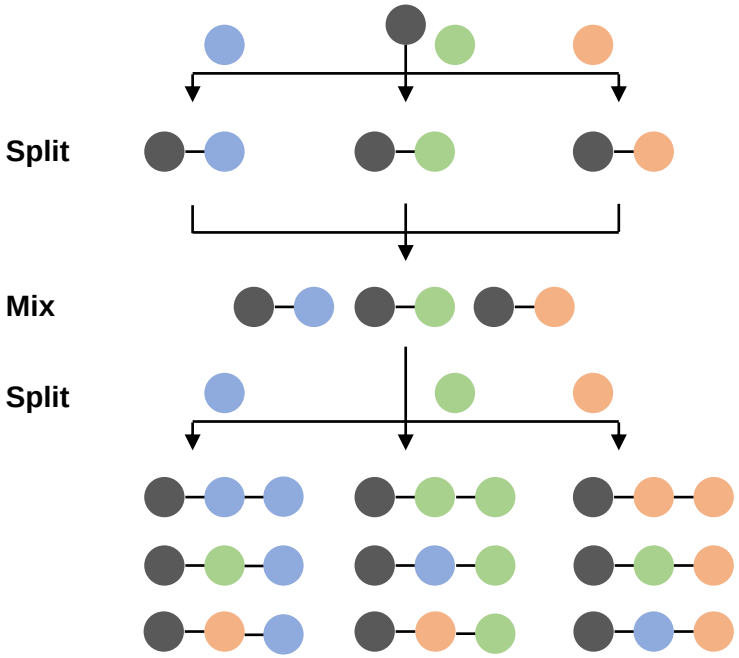


Figure 2.1: Diagram illustrating the split and mix synthesis concept to generate a one-bead one-compound library (grey circles are micro beads). Here, 3 building blocks (blue, green and orange circles) are used to generate a didipeptide library which gives $3^2 = 9$ compounds.[56]

Post synthesis, there are two approaches to screening the libraries: solution-based or directly on-bead. The solution-based approach is most aptly demonstrated by Schreiber *et al* who demonstrated their 'one-bead, one-stock solution' technology platform.[57] This approach involved synthesis of small molecules on macrobeads which delivered stock solutions of compound after cleavage and resuspension in assay well plates.[58] These can then be screened in phenotypic microarray assays. The drawbacks associated with the solution based approach post-synthesis are difficulties with compound identification, small quantities of compounds, low purity of pooled compounds and the requirement for re-synthesis of hit compounds.[59]

In the on-bead screening (OBS) approach, immobilised ligands are screened directly at the site of synthesis, this provides a number of advantages. For example, there is no need to purify compounds prior to screening, only hits need to be resynthesized and characterised. Also, identification of low affinity ligands, up to millimolar dissociation constants (K_D) can be detected.[60] The library is mixed with the target molecule, for example, protein, DNA or whole

cells, and interacting compounds can be identified by appropriate reporter systems, such as fluorescence or colourimetric enzymatic assays.[61] The primary method for on-bead screening in many academic groups, involves the use of a bead sorter (Union Biometrica, COPAS) and standard fluorescent microscope equipment. Hits are identified using a fluorescent target molecule and generally fluorescence imaging detects the signal over the entire bead volume. This introduces an issue with background fluorescence from the bead matrix and can lead to 'picking' of false positive beads.[62,63]Another drawback is that calculated K_D values from OBS were generally not reproducible in solution, and that unspecific binding was generally a concern.

2.1.3. The CTB On-Bead Screening Platform

The Auer lab (CTB) has extensive experience in on-bead screening methods for identifying new ligands from diverse one-bead one-compound (OBOC) chemical combinatorial libraries.[62,55,65–67] To address the issues of the COPAS detection method with bead autofluorescence, the establishment of confocal nanoscanning (CONA) provided the technology to focus on the outer circumference of microbeads where most of the binding events occur, as proteins are generally too large to diffuse into the matrix. This approach allows one to differentiate between real binding events and bead autofluorescence, hence minimising false positives. Also, the libraries are generally synthesised on TentaGel™ resin, made of a crosslinked polystyrene matrix onto which poly-ethylene glycol (PEG) is attached. The PEG groups are functionalised with an amine group for initiating peptide synthesis using solid-phase peptide synthesis (SPPS). High quality amino acid building blocks are widely available and many modified derivative and non-natural amino acids are also commercially available. TentaGel™ resin also exhibits minimal autofluorescence and increased mechanical stability, opposed to polystyrene-based resins, therefore is a big advantage. The beads, exhibiting the library compounds, are incubated with a fluorescently labelled target protein. Screening is then performed using a scanning confocal microscope for the detection of the bound protein via its fluorescence. The imaged obtained is a cross section of the beads in which binding events can be detected as halos or "rings" around the beads. Beads are picked using a highly automated bead picking instrument, PickoScreen-02 (PS02), developed in collaboration with Perkin Elmer, Germany. For hit compound deconvolution and K_D determination, a method for fluorescently labelling the active compound was devised, this was named post synthesis/post screening labelling (PS/PS). A small tagging group, is

incorporated into the structure of the screening compounds for attaching a dye following screening, this reduces the risk of non-specific interactions and provides a means for K_D determination using techniques such as fluorescence correlation spectroscopy (FCS). The tagging group commonly used in the Auer group is a terminal alkyne containing amino acid (propargyl glycine) which can be labelled using the copper catalysed Huisgen 1,3-dipolar addition 'click' reaction between an alkyne and an azide functionalised dye before cleavage. The Pra residue is stable to the chemical conditions of peptide synthesis so is useful for biorthogonal chemistry. Azide functionalised tetramethylrhodamine (TMR) is the frequently employed dye used due to its excellent fluorescence properties and stability to peptide cleavage conditions (see figure 2.3).

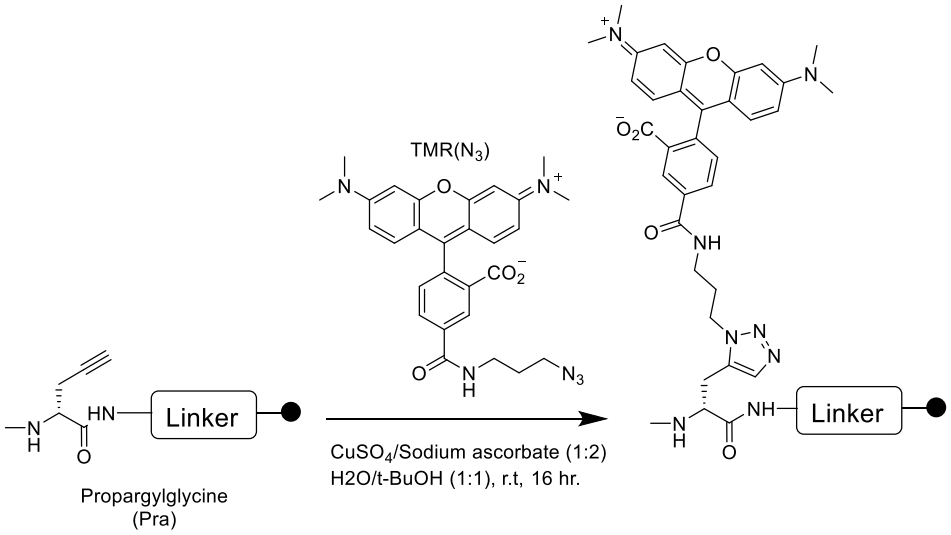


Figure 2.2: Scheme of click chemistry for single bead labelling. A terminal alkyne bearing amino acid (Pra) is incorporated into the library construct. After hit bead detection, copper catalysed cycloaddition allows labelling using an N_3 functionalised dye, for example tetramethylrhodamine azide (TMR(N_3)).

The use of a cleavable linker is required for on bead screening for deconvolution of hit compounds after screening, the linker used in this project is a modified version of SCAL (Safety Catch Acid Labile) linker.[68] Primarily, the linker is required to allow for orthogonal side chain deprotection, for screening, and subsequent peptide cleavage, this is illustrated in figure 2.3. Also, a space group is attached to increase the distance between the library compound and the label, minimising the effect of the label on the binding affinity. Finally, the peptide is built up by stepwise addition of Fmoc-protected amino acids, a schematic of the bead components is illustrated in figure 2.3.

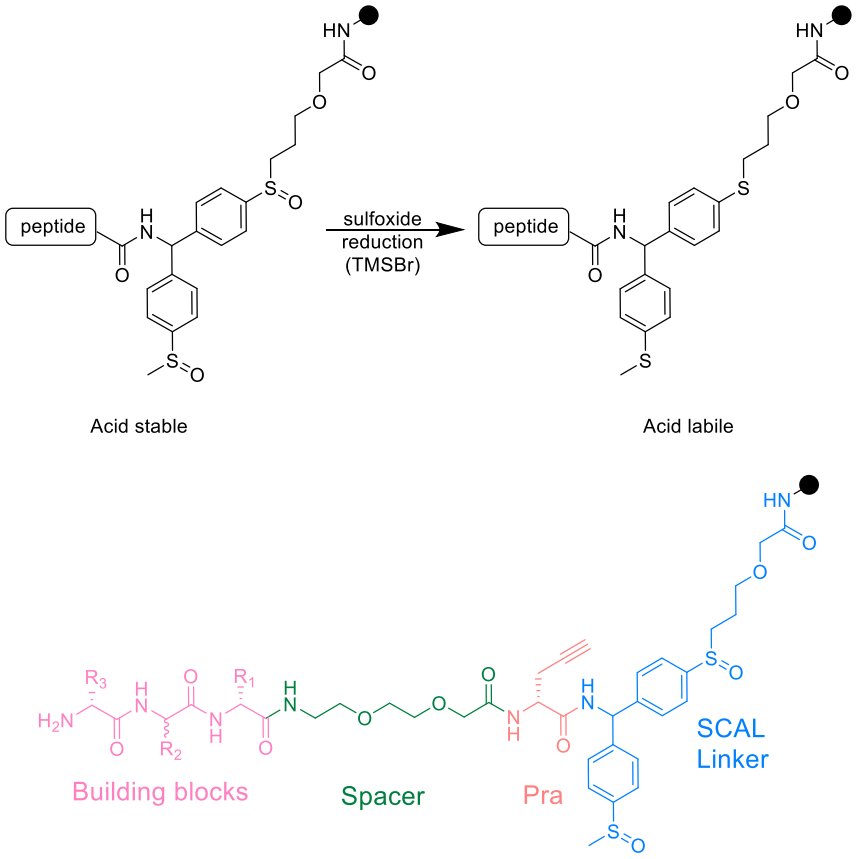


Figure 2.3: Top: Schematic of SCAL2 linker. The peptide is synthesised by standard SPPS methods using protected amino acids. The protection groups can then be cleaved by treatment with TFA, leaving the native peptide still attached to the bead and therefore available for screening. Reduction of the sulfoxide groups by treatment with TMSBr allows the native peptide to be released from the resin for solution binding studies. Bottom: Diagram of the components for solid-phase synthesis of peptide libraries. TentaGel™ beads (black circle), SCAL linker (blue), Pra (propargyl glycine, labelling site, red), Spacer (green), example of trimpeptide library building blocks (magenta, R_1, R_2, R_3 are random amino acid side chains).

Post screening, hit compounds are cleaved and usually deconvoluted by mass spectrometry for identification. This highly automatable and high throughput screening platform has been proven to work reliably in the synthesis of > 60 small molecule and peptidomimetic libraries for many PPI targets including: Mdm2, HuR, Importin-beta, LFA-1 and SAP, to name a few.[65,67,69]

2.1.4. The Tripeptide Concept

A popular metric used in drug discovery to narrow the search for lead compounds is ligand efficiency (LE). This is defined as the binding free energy of a ligand divided by its molecular size (i.e. number of heavy atoms (HA)).[70] It is frequently presumed that there is a level of additivity between different functional groups of a given ligand. This assumes that the overall

affinity of a ligand is a linear combination of the affinities of the constituent pieces.[71,72] However, Reynolds *et al* re-examined the extensive amount of literature on protein-ligand binding affinities and show that up to a certain limit of heavy atoms (or molecular size) the maximal affinity is linear and then plateaus (this is illustrated in figure 2.4). The overall trend is that for a given number of HAs the change in affinity is not linear with size and is observed across a wide range of protein-ligand interactions.

In theory, an OBOC peptide library can contain peptides of any length, however, synthesis beyond ~ 20 amino acids tends to generate problems. The concept behind the use of tripeptides for screening is that the smallest information for specific recognition events between amino acids in endogenous peptidic ligands and their native receptors are three amino acids. It has been shown repeatedly that tripeptides have important signalling roles in biology and are capable of various functions.[73,74] For example, Rajasekhar *et al* have developed a multifunctional peptidomimetic inhibitor of amyloid beta (Aβ) aggregates, a hallmark of Alzheimer's disease. This was based on conjugation of a natural, metal chelating tripeptide (GHK) and a peptidomimetic analogue of the core recognition unit (Sr-V-Sr-F-Sr, Sr: sarcosine) of Aβ fibrils.[75] The modified peptide interacts with Aβ monomers and prevents formation of toxic Aβ oligomeric and fibrillar aggregates. This parallel peptide/peptidomimetic approach excellently demonstrates the use of natural peptides and peptidomimetic tripeptides as potential therapeutic agents. Also, one can apply the tripeptide concept to the ligand efficiency metric described by Reynolds. Figure 2.4 shows the average number of HAs in a single amino acid, a dipeptide and a tripeptide as red, yellow and green lines respectively. This shows that a tripeptide should theoretically be at the maximal affinity obtainable for ligand efficiency. Identification of a tripeptide or peptidomimetic would allow development of a tool compounds to study protein activity in a variety of assays.

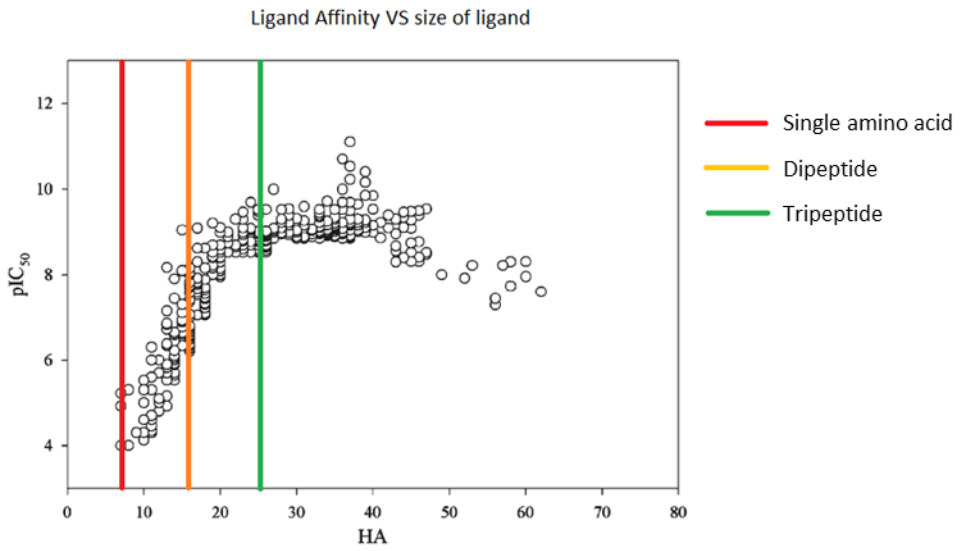


Figure 2.4: The 'maximum affinities' of compounds (as measured by IC_{50}) increases rapidly up to 20 heavy atoms (HA; non-hydrogen atoms), but plateaus beyond 25. For reference, the average number of heavy atoms in a single amino acid, dipeptide and tripeptide are shown as red, yellow and green lines respectively. Figure adapted from Reynolds *et al*.[76]

2.1.5. Fluorinated Peptides and NMR Screening

On-bead screening and solution-based screening can be very efficient, however, for short peptides which might further be converted into drug like molecules by medicinal chemistry, for example, tripeptides with the highest ligand binding efficiencies, a method is required which works with lower affinity to allow for higher μM K_Ds.

NMR spectroscopy is a well-established, solution-based method for detecting ligand binding, however, literature is relatively scarce on the screening of peptide libraries. As described earlier (Chapter 1.4), NMR-based screening approaches are generally divided into two categories: ligand-observed and protein-observed. NMR screening of peptide libraries has generally been reserved to protein-observed methods owing to higher sensitivity at weaker affinities.[77] However, the disadvantage is that hit compound deconvolution is difficult as the ligands are not observed, therefore compounds need to be individually retested. Ligand-observed screening therefore offers a host of advantages. SPPS allows the incorporation of non-natural amino acids and using a fluorinated amino acid derivative in a peptide library alleviates the issue of compound deconvolution. This is because the ^{19}F nucleus is highly sensitive and peptides constructed using a fluorinated amino acid will show a unique fluorine resonance. This is in contrast to ^{1}H observed methods where, again, hit compound deconvolution is difficult due to spectral overlap of proton ligand resonances. Fluorine NMR

screening is described in detail in chapter 1. An example of the broad applicability of fluorinated peptides comes from Howard and co-workers who describe a combined approach of ligand-observe ^{19}F and protein-observe ^{15}N-HSQC NMR to study peptide binding to the human protein disulfide isomerase (hPDI). The peptide-ligand is a 14 *mer* (H_2N-AGSKNFFWKTFTSS-OH) that binds to hPDI with K_D ~ 100 μM and stepwise mutations to the phenylalanine residues increased the K_D until all binding was abolished. They also highlight the advantage that ^{19}F offers a more accurate measure of K_D due to higher spectrum resolution.[78]

The generation of peptide libraries is generally limited by the stipulation for chemical or biological tags to allow easy identification. Similarly biological approaches to peptide library synthesis, for example, phage display rely on multiple rounds of biopanning and sequencing to identify the highest affinity peptides.[79,80] Therefore, the use of ^{19}F as a sensor for ligand binding detection in a library of fluorinated peptides would be a simple approach for low affinity binding detection and simple hit deconvolution.

2.1.6. Target Selection and Motivation

As described earlier (Chapter 1.2), target selection and validation is the first stage of the drug discovery pipeline. Moreover, the majority of the ~ 25,000 human genes have not been fully characterised for their implication in disease.[81] The common definition of a validated drug target is that up- or down regulation and/or knockout models *in vivo* produces a disease relevant phenotype. However, generating a knockout model is time-consuming and upregulation is not always straightforward, this is biological target validation. Another method to identify a potential target, is to develop a known, low molecular weight ligand to use as a tool compound for studying molecular effects and concentration dependent modulation of inhibition. This is chemical validation and addresses the specific ligand binding sites and, hence, potential for inhibition. The target focus of this thesis are PPIs and discovery of inhibitors/modulators of these interactions for studying their therapeutic benefit. The following sections introduce two important target proteins used in this project.

2.1.7. ATPase AAA Domain Containing 2 (ATAD2)

The epigenome is a functional record of chemical modifications to the DNA and histone proteins that moderate gene regulation. Currently, there are over 16 major histone chemical modifications that determine the transcriptional status of a cell, including: methylation,

acetylation, phosphorylation and ubiquitination.[82] Epigenetic regulators are the main proteins involved in histone remodelling/modifying and are involved in a myriad of PPIs governing cellular dynamics. Histone acetylation is one of the most studied histone modifications that implicates an open chromatin structure and hence involved in moderating DNA transcription. The acetylation of lysine residues within the N-terminal domain of histone tails is found across the entire epigenome. There are three main families of proteins dictating the levels of acetylation on histone tails, these are: histone acetyl transferases (HATs), histone deacetylases (HDACs) and bromodomains (BRDs). HATs catalyse the transfer of an acetyl group to the ε-amino group of lysine side chain residues in the N-terminal chain of histone tails, whereas HDACs' function is to remove it. BRDs are the primary protein modules that recognise the acetylated epigenetic mark.[83]

BRDs are a family of evolutionarily conserved protein interaction modules and the human proteome encodes 61 BRDs present in 46 diverse nuclear and cytoplasmic proteins. Despite the surprisingly large sequence variations, BRDs share a conserved tertiary structure that comprises of 4 α-helices (αZ, αA, αB and αC) linked by diverse loop regions of variable length and charge. The acetyl lysine binding site is situated at one end of the helical bundle in a centrally located hydrophobic pocket (see figure 2.5).[84] The acetyl lysine is recognised by a conserved asparagine residue mediated by hydrogen bonds from conserved water molecules. It has been shown that some BRD modules (i.e. BET subclass of proteins) simultaneously recognise two acetylated lysine histone marks on the same histone tail.

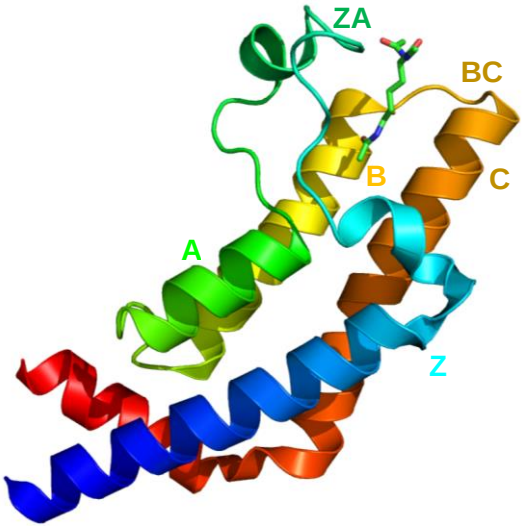


Figure 2.5: Structure of the BRD of ATAD2 in complex with capped acetyllysine (PDB ID: 5A5N). The structure is a conserved bromodomain 4-α-helical bundle (A, B, C and Z) linked

by two diverse loop regions (ZA and BC). The acetyllysine binding site is located between the ZA and BC loops. [85]

ATAD2 (ATPase family AAA Domain containing 2), also called ANCCA (AAA Nuclear Co-regulator Cancer Associated), is a transcriptional regulator that carries two AAA (ATPases Associated with diverse cellular Activites) domains and a single BRD. The overexpression of human ATAD2 has been associated with many cancer types and Zou *et al* have shown that ATAD2 is strongly induced by estrogen in human breast cancer and is required for estrogen induced cell proliferation.[86] Several other reviews of this interesting protein have exposed its role as a genetic marker for poor prognosis of cancer suffering patients.[87–91]

Due to the growing significance of transcriptional dysregulation in disease, BRDs have emerged as a potential target for pharmacological intervention. Knapp *et al.* have recently reported highly potent small molecule inhibitors of the BET (Bromodomain and Extra-Terminal containing) family of proteins largely due to the hydrophobic nature of the acetyl lysine binding pocket.[92] Current drug discovery efforts are focused largely on the BET family of bromodomain containing proteins, with a large volume of literature originating in this family of proteins.[93–95] The proven druggability of bromodomains, together with their central role in tumorigenesis and disease, suggests an appealing target for drug intervention.[96]

2.1.8. Ubiquitin Conjugating Enzyme L3 (UbE2L3)

The ubiquitin proteasome system (UPS) is the principal network for protein catabolism which is central to the regulation of almost all cellular processes. Ubiquitination is a proprietary mechanism of the UPS through which transfer of a ubiquitin molecule to a substrate protein occurs via thioester transfer by way of an E1-E2-E3 enzymatic cascade (figure 2.6). The resultant mono- or poly-ubiquitinated substrate is then typically targeted for degradation by the 26S proteasome, although ubiquitination also plays a role in other cellular processes such as: cell signalling and regulation of DNA transcription.[97,98] The prime interactor in this enzymatic cascade is the E2 conjugating enzyme which performs a variety of functional roles. The E2 is first charged with ubiquitin on its active site cysteine via an E1-catalysed ATP-dependent thioesterification. The E2 is then responsible for a subsequent transfer of ubiquitin to nucleophilic reactive residues on substrate proteins *via*: transthiolation (transfer to Cys), esterification (transfer to Ser/Thr), amide formation (transfer to N-terminus) and isopeptide formation (transfer to Lys residues). The final step to substrate ubiquitination often occurs with the aid of an E3 ligase which regulates substrate specificity.

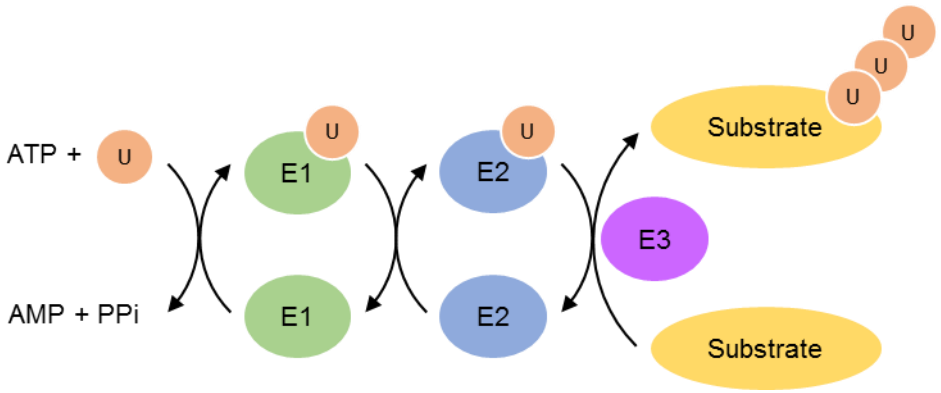


Figure 2.6: Scheme illustrating a typical ubiquitination reaction. The activation of ubiquitin occurs *via* an ATP dependent thioesterification transfer to an E1 enzyme. A transthiolation reaction transfers the ubiquitin to an E2 enzyme. There are then two mechanisms by which substrate ubiquitination can occur: 1) ubiquitin transfer occurs directly from E2 to substrate *via* a PPI mediated by an E3 enzyme or 2) ubiquitin is transferred *via* another transthiolation to the E3 which then directly ubiquitinates the substrate.[99]

Due to the importance of UPS in regulating cellular processes, defects in this system can result in pathogenesis of several important human diseases, including: autoimmune diseases and Parkinson's. Table 2.1 describes an overview of E2 implications in disease.

Name and synonyms	Cellular processes	Disease association
UbE2C;UbcH10	Cell cycle	Cancer[100]
UbE2D4; UbcH5d	Apoptosis[101]	
UbE2E2		Diabetes, lung cancer[102]
UbE2K;HIP2;UBC1; E2-25K	Apoptosis, proliferation	Parkinson's, Alzheimer's and Huntington's diseases, cervical cancer[103–106]
UbE2L3/UbcH7	Cell cycle	Autoimmune diseases, Parkinson's disease[107,108]
UbE2L6/UbcH8	Immune response	Cancer, obesity, Parkinson's disease, viral infection[109–111]
UbE2Q2	Cell cycle, apoptosis	Cancer, renal pathologies[112,113]

Table 2.1: A variety of E2 enzymes and their cellular processes with their implication in diseases. Adapted from Koszela.[114]

Specifically, UbE2L3 is one of the 38 E2 enzymes encoded by the human genome. Several genome-wide association studies (GWAS) have identified various polymorphisms in the

genomic locus of UbE2L3 associated with autoimmune and neurodegenerative diseases.[107] A revolutionary discovery identified it as one of the few E2s that only exhibits transfer activity towards cysteine residues on E3 ligases.[115] Recognising its implication in disease lends itself as a potential target for drug intervention, and as of 2018 there are no known specific inhibitors of this particular E2. The multifunctional role of the E2 proteins in PPIs with E1, E3 and ubiquitin provide multiple opportunities for modification of E2 activity. That is, modification of activity can be addressed by targeting the E1 binding site or E3 binding sites, in addition to the conventional inhibition of the E2 catalytic site. Also, allosteric moderators of this E2 enzyme may interfere with the PPIs governing interactions with both E1 and E3 enzymes, for example the small molecule inhibitor CC0651 for UbE2R1.[116]An example of an E3 ligase associated with UbE2L3 activity is Parkin, of which mutations are associated with mitochondrial dysfunction leading to neuronal damage in Parkinson's disease.[108,117] Hence, direct binding site inhibition of this E2 will prevent the ubiquitination cascade and prevent premature neuronal cell death.

2.1.9. Aims

The aim of this part of the thesis work was to explore new concepts of how short peptides and peptidomimetics (mainly di- and tripeptides) could be screened against PPI targets. There are two approaches presented here for the use of peptides as screening agents.

I. The first utilises the OBOC approach to synthesise a bromodomain-focussed tripeptide library that can be used for screening ATAD2. This BRD has a large, open hydrophobic binding pocket and it is hypothesised that using a non-natural lysine derivative at the central position of each tripeptide in the library will uncover a unique binding partner. Also, it was necessary to produce a positive control peptide for binding evaluation. Therefore, an additional aim was to synthesise a known peptide pentapeptide mimic of the histone H4 tail and evaluate it as a tool compound for binding against ATAD2 using various NMR methods.

II. The second approach introduces an 'on-protein peptide growing' (OPPG) method by using fluorinated amino acids as chemical probes for binding detection. The hypothesis was that using a fluorinated amino acid in an initial dipeptide library would provide a sensitive probe to detect weak affinity dipeptides by ^{19}F NMR. Hit dipeptides could then be extended, either C- or N-terminally with natural or non-coding amino acids, to increase the binding affinity to the target protein. This idea is

outlined in figure 2.7. It was envisaged to perform a proof of concept study to determine the feasibility of the approach. For this we applied a known tetrapeptide peptidomimetic, which had been developed in the Auer lab, towards the inhibitor of apoptosis (IAP) related protein, survivin.[118] The final, ratified OPPG approach was applied to the UbE2L3 protein target.

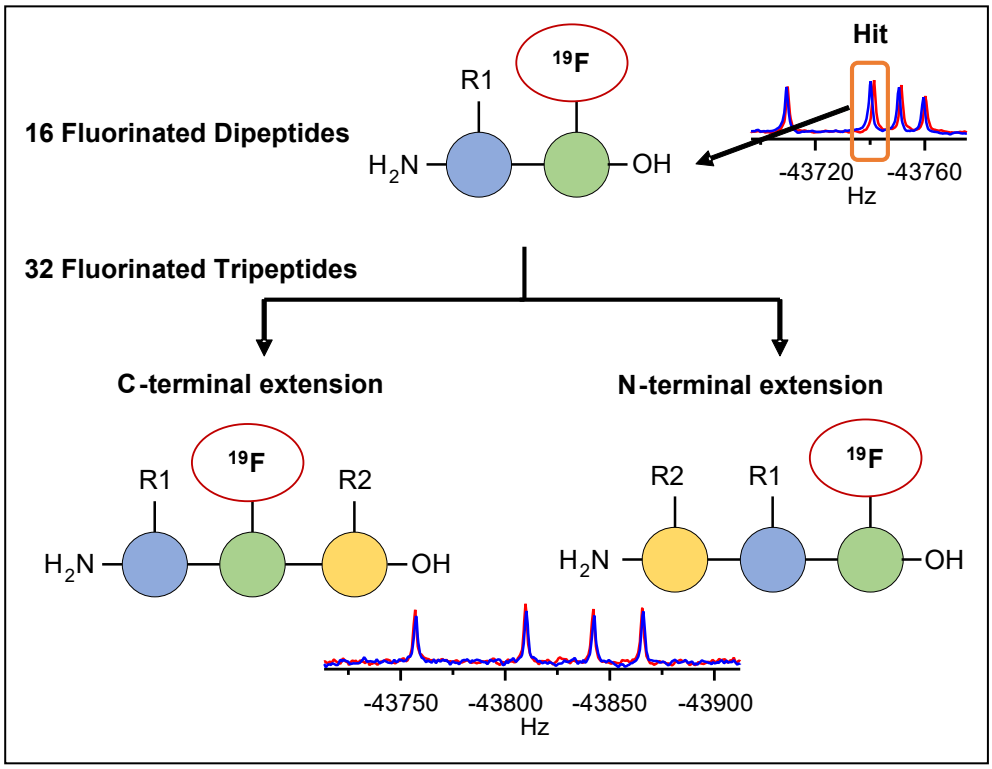


Figure 2.7: Schematic of the 'on-protein peptide growing' approach using a fluorinated amino acid as a probe molecule for binding detection. The initial dipeptide library is screened by ^{19}F NMR and hits can be deconvoluted based on changes in ^{19}F resonance signals. A hit dipeptide can be extended to a tripeptide by amino acid addition at both the 'N' or 'C' terminus and binding measured again by ^{19}F NMR. R1 and R2 indicate randomised natural amino acid residues. ^{19}F is the fluorinated amino acid probe.

2.2. Results and Discussion

2.2.1. Peptide Tools for Studying ATAD2

The Structural Genomics Consortium (SGC) is a public-private partnership with the core directive to determine 3D structures of human proteins of clinical importance to accelerate the drug discovery process. In collaboration the Knapp lab, at the SGC University of Oxford, supplied structural data for the ATAD2 BRD protein in complex with a pentapeptide, herein named H4K5. H4K5 is a peptide mimic of histone H4, with an acetylated (Ac) lysine residue at position K5 (figure 2.8).[119]

1. H_2N-MSGRGKGGKGLGKGGAKRHR……
2. H_2N-MSGRGKGGKGLGKGGAKRHR……

Figure 2.8: The N-terminal sequence of histone-H4 (shown N→C). 1. The highlighted region is the pentapeptide that was used for crystal soaking at the SGC. The protein sequence starts at M0, and the pentapeptide is R3 – G7. The lysine at position 5 was acetylated for the soaking experiments. 2. The highlighted region here is peptide H4 (figure 2.13), sequence from R3 – G13.

Examining the structure of the H4K5 peptide in complex with ATAD2 (figure 2.9), it can be seen that the C-terminus is extended into space and does not appear to contribute any significant binding contacts. It was therefore hypothesised that addition of a fluorinated amino acid probe at the C-terminus separated by a small spacer would not hinder the binding effect and provide a useful probe for binding detection of by ^{19}F NMR.

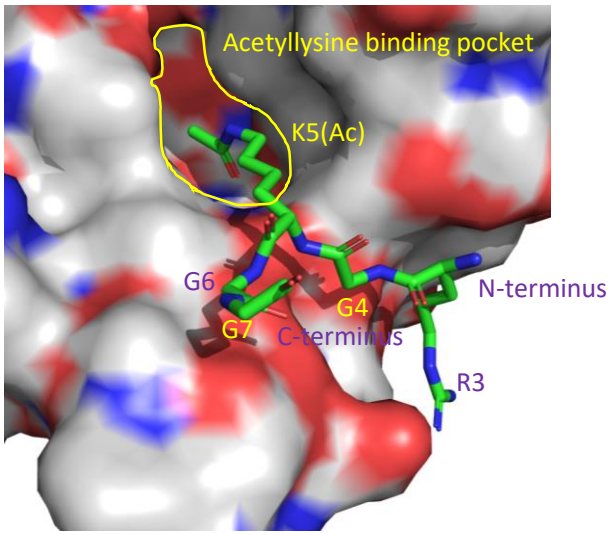


Figure 2.9: Structure of the ATAD2 binding site in complex with peptide H4K5(Ac) (green) (PDB ID: 4QUU).[119] Peptide sequence: H_2N-RGK(Ac)GG-OH (R3-G7).

2.2.2. ATAD2 expression, purification and ^{15}N labelling

ATAD2 was successfully expressed in *E. coli* (expression strain Rosetta (DE3) pLysS) using traditional bacterial shake flask expression methods. The protein was purified by gravity flow using immobilized metal ion affinity chromatography on Ni-NTA agarose resin. The yield from a 1L bacterial cell culture was excellent: 19 $mg.L^{-1}$ of culture and purity was measured at > 95% by analytical HPLC and SDS-PAGE. For ^{15}N-HSQC NMR experiments, isotopic labelling with ^{15}N was required. This was achieved by traditional bacterial shake flask expression in minimal media and ^{15}N ammonium sulfate ($(^{15}NH_4)_2SO_4$) as the sole nitrogen source. Again, the protein was purified by gravity flow using immobilized metal ion affinity chromatography on Ni-NTA agarose resin. The yield from a 1L bacterial cell culture was good: 4 $mg.L^{-1}$ of culture and purity was measured at > 95% by analytical HPLC and SDS-PAGE (figure 2.10, see chapter 7.2.1 for full materials and methods and analysis).

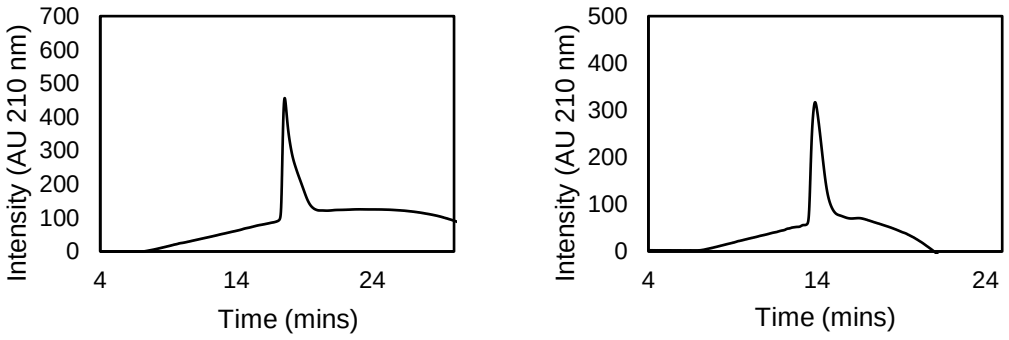


Figure 2.10: HPLC chromatogram of, left: ATAD2, t_R 15.6 mins. Right: ^{15}N-ATAD2 t_R 14.0 mins. (N.B. gradient methods were different, hence different retention times, see Chapter 7 for details).

2.2.3. Synthesis of Fmoc-acetyl lysine

The acetylated lysine is the central amino acid for the H4K5 peptide. Fmoc protected acetyl lysine can be synthesised in the lab using available reagents (see figure 2.11). The materials and methods for synthesis of Fmoc-*N*-acetyl-lysine-OH are given in chapter 7. The final compound was analysed by HPLC and showed a purity > 95% and LCMS shows the correct ion peak: $[M+H]^+_{calc}$ = 411.2, $[M+H]^+_{obs}$ = 411.1. ^1H and ^{13}C NMR spectra were also acquired to determine correct structure. Figure 2.11 shows the HPLC chromatogram of the final compound. See appendix 8.1.4 for mass spectrum analysis.

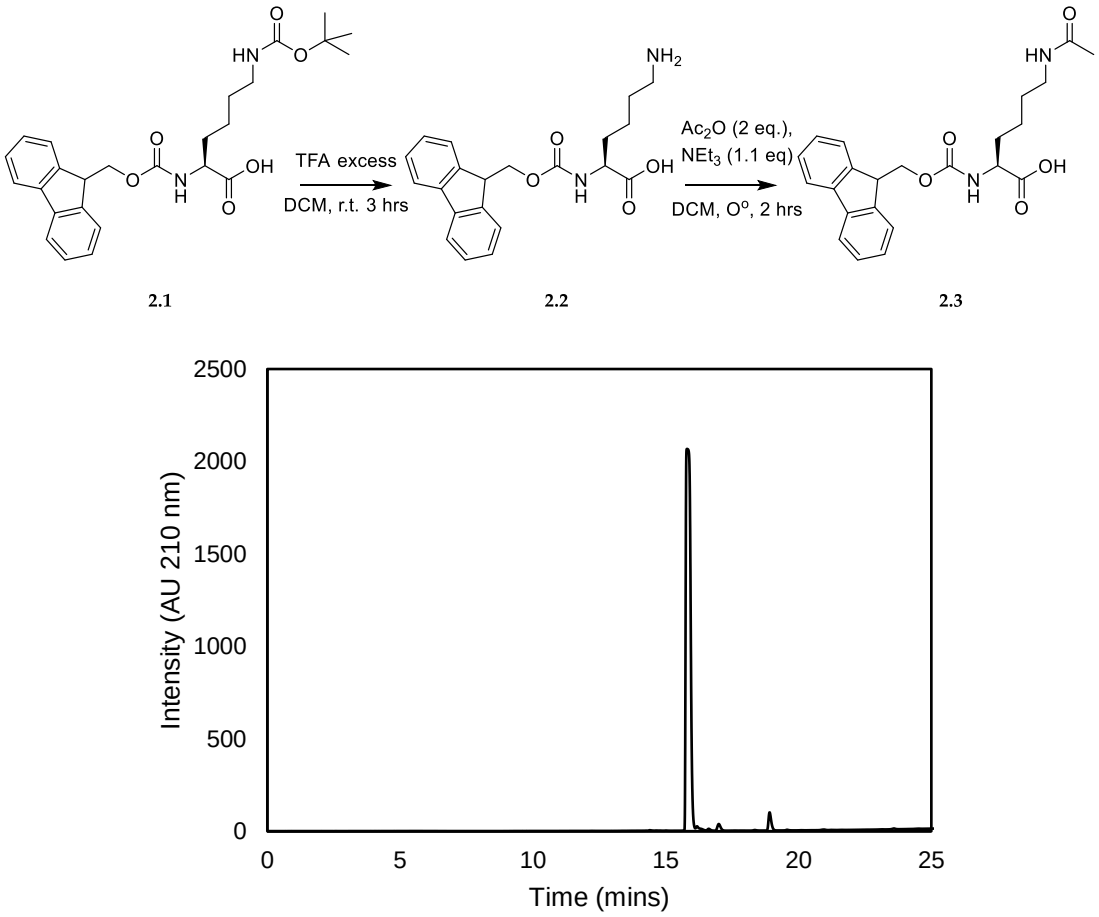


Figure 2.11: Top: Scheme for synthesis of Fmoc acetyl lysine. The Boc protected ε-amino of Fmoc-Lysine(boc)-OH (**2.1)** is first deprotected to Fmoc-Lysine-OH (**2.2**). Then acetylated using acetic anhydride to generate Fmoc-*N*-acetyl-L-Lysine-OH (**2.3**). Bottom: HPLC chromatogram of N_α-Fmoc-N_ε-acetyl-*L*-Lysine-OH. t_R = 15.9 mins; purity > 95 % as measured by peak area.

2.2.4. ATAD2 peptide synthesis and binding detection by ^{19}F NMR

The peptides synthesised for use in this section are illustrated in figures 2.12 and 2.13. All peptides were synthesised by Fmoc SPPS and identified by LCMS and purity measured by HPLC (see chapter 7.2.9 for materials and methods and appendix 8.1 for QC analytics).

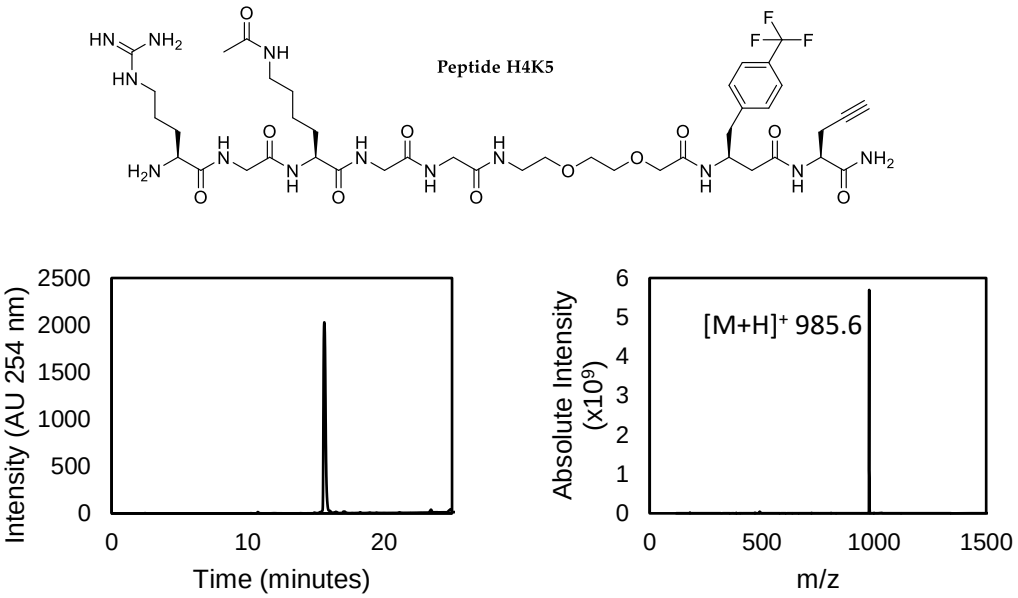


Figure 2.12: Top: chemical structure of peptide H4K5. Bottom: HPLC chromatogram and mass spectrum of peptide H4K5. HPLC: t_R = 15.6 mins. Absorbance is measured at 254 nm, and purity is > 95% (as measured by peak area). LCMS mass spectrum of peptide H4K5 (PRA-β(*m*CF_3)F-DOA-G-G-K_{ac}-G-R). $[M+H]^+_{calc}$ 985.4; $[M+H]^+_{obs}$ 985.6.

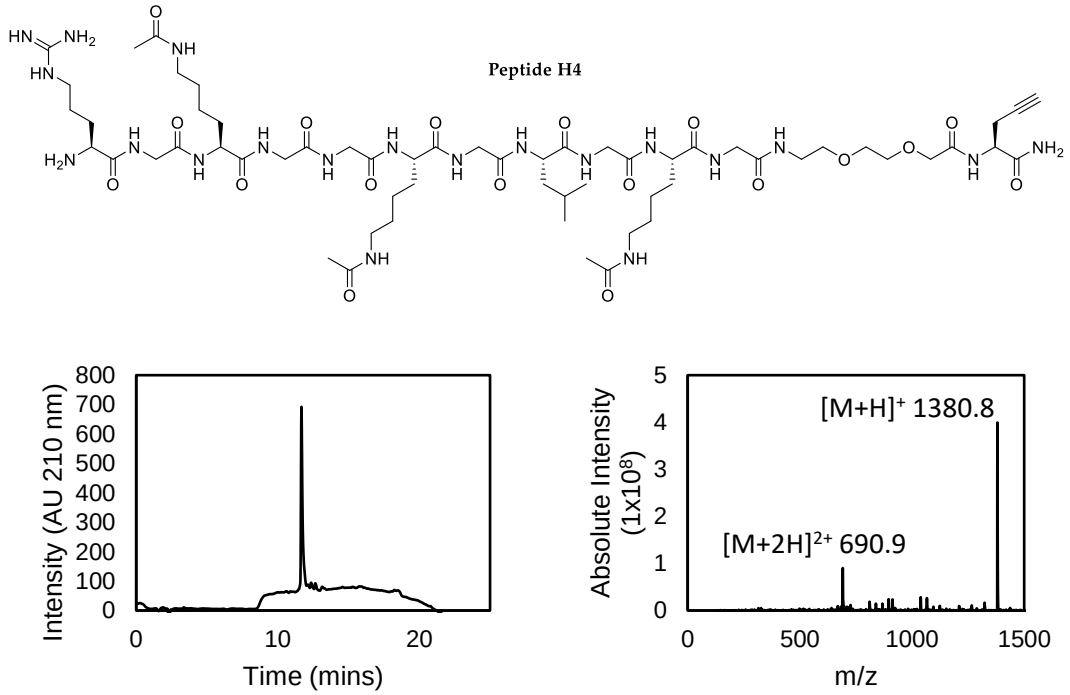


Figure 2.13: Top: chemical structure of peptide H4. Bottom: HPLC chromatogram and mass spectrum of peptide H4. HPLC H4. t_R = 11.7 mins. Absorbance is measured at 210 nm, and purity is > 90% (as measured by peak area). LCMS mass spectrum of peptide H4: $C_{59}H_{102}N_{20}O_{18}$ (H_2N-RGK_{ac}GK_{ac}GLGK_{ac}G-DOA-PRA-OH). $[M+H]^+_{calc}$ 1378.8; $[M+H]^+_{obs}$ 1380.8.

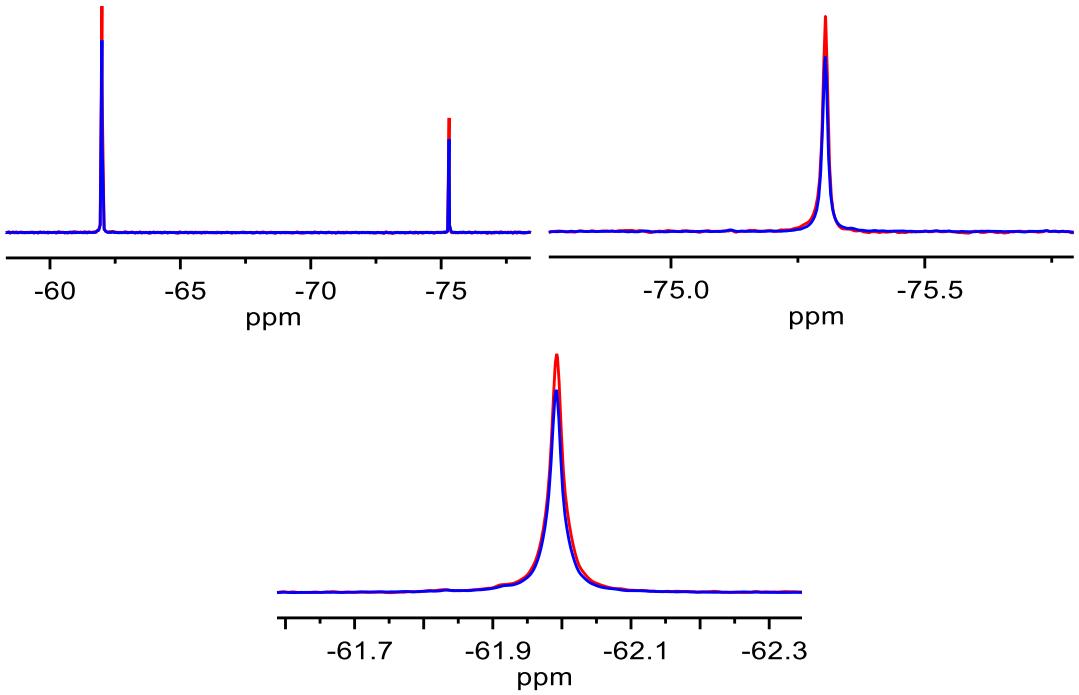


Figure 2.14: A superposition of 19F NMR spectra of fluorinated peptide H4K5 (1.3 mM) in the absence (red) and presence (blue) of 200 μM ATAD2 (1:6.5 protein:ligand molar ratio). Top: full spectrum. Middle: expanded region of the peak at -75.3 ppm corresponding to TFA. Bottom: expanded region of the peak at -62 ppm corresponding to the fluorinated peptide. On the addition of protein there is no visible change in the chemical shift or line width of the fluorine resonance, indicating no binding between the ligand and protein. Spectra are analysed using MestReNova v10.0.2-15465 (see Chapter 7 for experimental details).

The fluorinated peptide H4K5 was used in a ^{19}F NMR binding assay against the ATAD2 bromodomain. A 1D ^{19}F NMR spectrum of the peptide was recorded in the presence and absence of protein (figure 2.14). The spectra are then easily analysed by superposition to identify spectral changes in the fluorine resonance. From figure 2.14, it can be seen that there is no distinct change in the fluorine resonance on addition of protein (blue spectrum). Peptide cleavage conditions require the use of TFA to remove peptides from the resin. The peptides are then precipitated in diethyl ether leaving the final peptide as a TFA salt. The TFA can therefore be seen in the ^{19}F NMR spectrum as a single peak at -75 ppm. This is commonly used as a reference peak, as TFA is a small organic molecule, it is unlikely to bind to the protein, and therefore any changes in chemical shift from the peptide can be referenced to the TFA peak. Using ^{19}F resonance as a sensor to detect binding is useful as there are no peaks from endogenous fluorine present in the buffer or from the protein, therefore, spectra are easily interpreted. However, there was no indication of the peptide binding to the protein in solution. This could be due to the addition of the spacer and/or fluorinated amino acid that

are sterically hindering the binding of the peptide. It could also just indicate that the fluorine nucleus is too far away in space to have a direct interaction with the protein. Although extremely sensitive to environmental changes, the fluorine nucleus may simply be only surrounded by solvent molecules and not directly interacting with the protein.

2.2.5. ^{15}N Labelling and Binding Studies of H4 Peptide Mimic to ATAD2

The premise behind ^{15}N isotopic labelling was to provide a higher sensitive method for binding detection of the native ligand. It was also hypothesised that a longer synthetic peptide mimicking the H4 histone tail would provide a higher binding affinity and, hence, a better chance of observing binding (structure of this peptide is given in figure 2.13). The peptide was also tri-acetylated according to information gleaned from the collaboration at the SGC. From herein this peptide is referred to as H4. This was aimed at developing a tool compound to study effects of inhibition, and provide a ligand for competition binding experiments.

The ^{15}N labelled ATAD2 bromodomain was initially analysed by 15N-HSQC NMR to determine the degree of labelling and the extent of folding. The ^{15}N-HSQC experiment provides a fingerprint of the backbone and side-chain amides of the protein and a correctly folded protein will show good peak dispersion of amide resonances. However, the HSQC spectrum of the protein does not show many resonances in the amide region and we would expect the number of resonances to equal the number of amide groups in the backbone. For the ATAD2 protein there are 145 amino acids (excluding prolines) therefore one would expect roughly 145 peaks in the spectrum. However, there are approximately only 60 distinct resonances in the amide region of the HSQC spectrum. This could indicate a misfolded protein or aggregation in solution. The protein itself required a high amount of salt (NaCl - 500 mM) in the buffer to solubilise it. The drawback with high salt concentrations is two-fold: 1) high salt concentrations can lead to protein aggregation in solution and 2) a high ionic strength buffer lessens the benefit of the CryoProbe™ and, hence makes it less sensitive. This could indicate the low signal intensities and number of amide resonances. A solution would be to increase the number of increments and number of scans per increment of the experiment to increase the resolution and signal to noise ratio respectively. Additionally, however, high salt concentrations are sometimes beneficial because this can indicate that binding is not driven by general nonspecific ionic interactions. Another problem encountered was that was difficult to concentrate the protein beyond 65 μM due to the high aggregation and precipitation observed.

The acetylated 11mer, H4 was used in a titration experiment with ^{15}N labelled ATAD2 BRD. A superposition of titration points is shown in figure 2.15. On addition of peptide H4, there are no perturbations to the backbone amide resonances in the spectrum. The titration molar ratios ranged from 1:1 to 1:16. Even at the highest concentration of ligand there are no changes in the backbone amide resonances, figure 2.16 shows a superposition of the first and last titration points for clarity. What is evident is that some signals disappear on the addition of H4, however, this could be due to a reduction in protein concentration during the titration as the final protein concentration is reduced by approximately 10%. However, signal disappearance is more likely due to unfolding of the protein as it appeared to be particularly unstable after being left at room temperature.

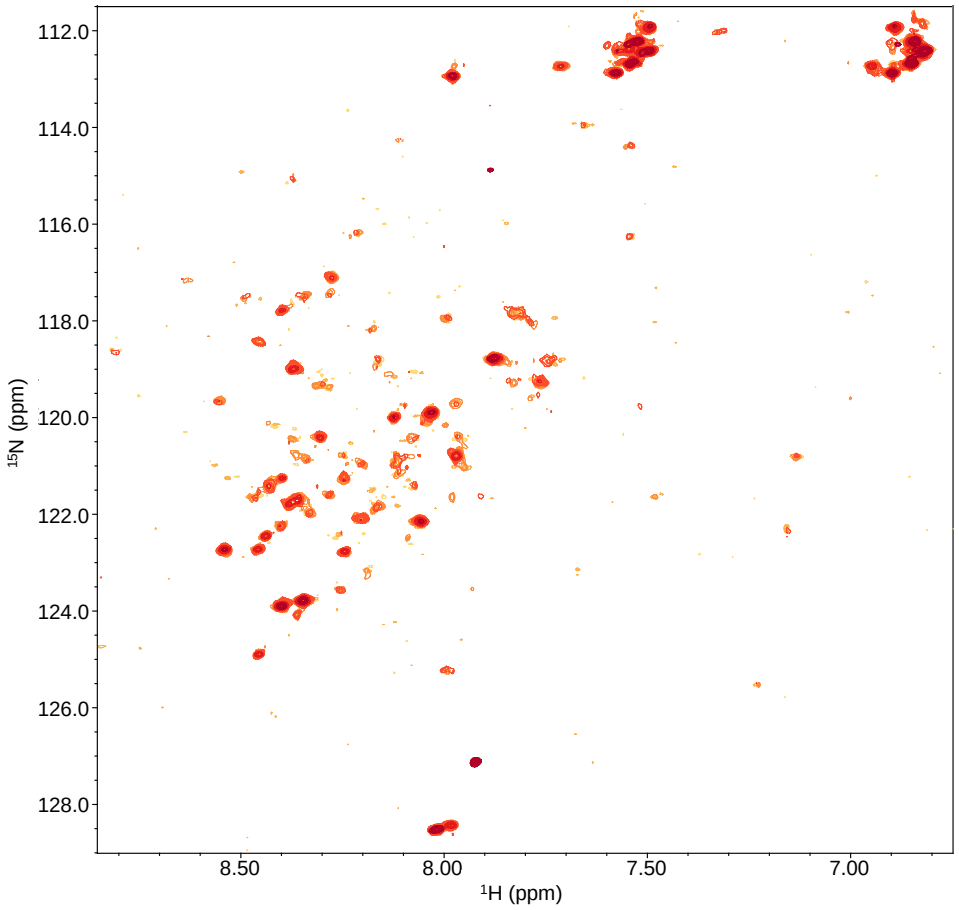


Figure 2.15: ^{15}N HSQC titration of ^{15}N-ATAD2 with peptide H4. From light to dark: 0, 50, 100, 200, 400, 1000 μM peptide and the protein concentration was 60 μM. The ^{15}N-HSQC experiments were performed on a Bruker Avance 800 MHz spectrometer in the School of Chemistry, University of Edinburgh. Each resonance signal corresponds to a backbone amide proton of the protein. However, some peaks are very weak and the resonance dispersion is not ideal. This could indicate an aggregating protein.

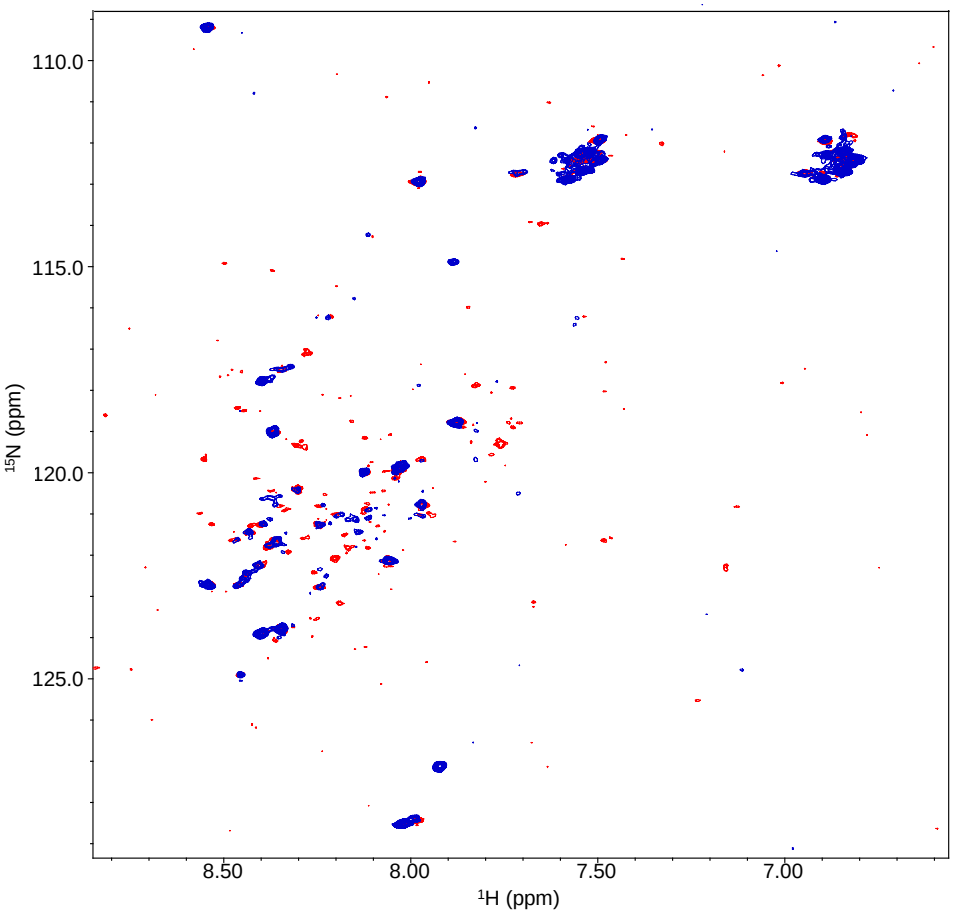


Figure 2.16: ^{15}N-HSQC titration of ^{15}N ATAD2 with peptide H4. This diagram shows a superposition of the first and last titration points of peptide H4 and ATAD2 BRD. The red spectrum is protein only, and the blue spectrum is the final peptide concentration of 1 mM and the protein concentration is 60 μM (1:17 protein:ligand molar ratio).

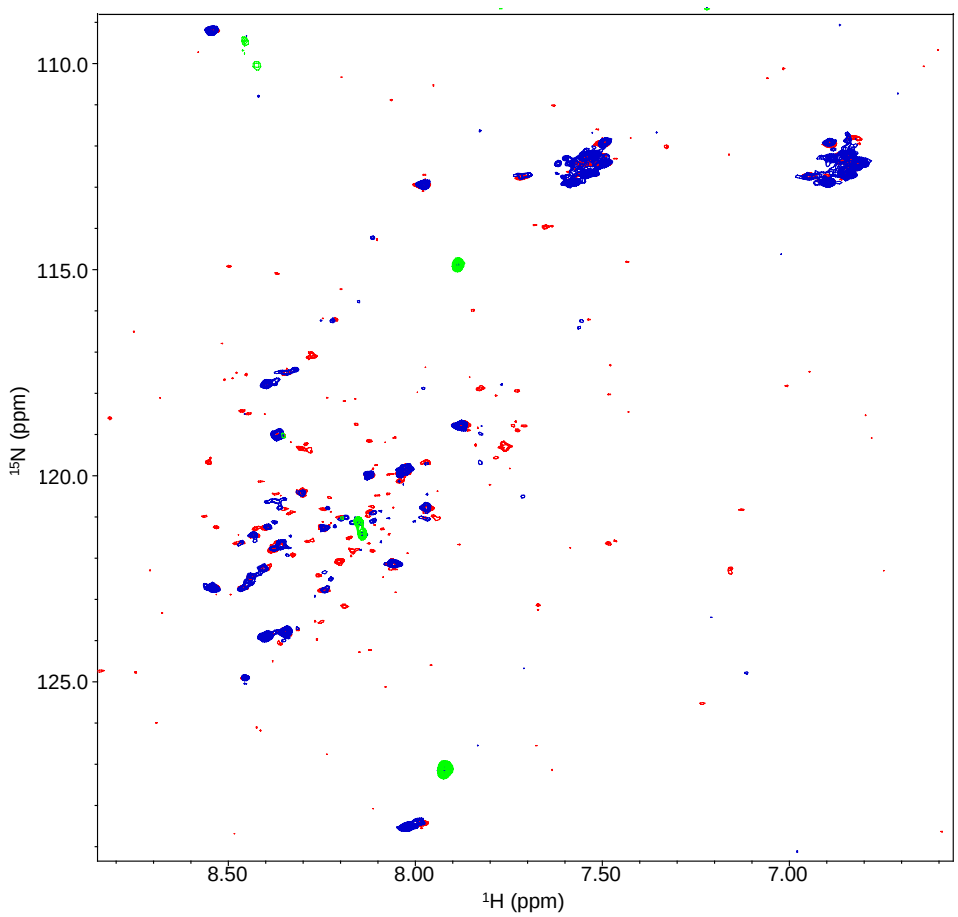


Figure 2.17: ^{15}N-HSQC titration of ^{15}N-ATAD2 with peptide H4. A superposition of the first and last titration points of peptide H4 and ATAD2 BRD. The red spectrum is the protein without peptide, and the blue spectrum is the protein with a final peptide concentration of 1 mM. The protein concentration is 60 μM (1:17 protein:ligand molar ratio). The green spectrum overlay also shows the ^{15}N natural abundance spectrum of the free peptide.

There are also some signals in the spectra that can be attributed to the natural abundance of ^{15}N in the H4 11*mer* peptide (figure 2.17). In summary, there is no evidence of ligand binding by ^{15}N-HSQC experiments with the native H4 peptide to the bromodomain of ATAD2.

2.2.6. OBOC Peptide Library Synthesis

In parallel to peptide tool compound development, a 'One Bead One Compound' tripeptide library was synthesized for on-bead screening against ATAD2 in search of a novel tripeptide ligand. The hypothesis is that ATAD2 has a very large, open, hydrophobic binding pocket. (figure 2.9), therefore extension of the lysine residue would provide a better capacity for non-

covalent binding interactions, for example, electrostatic interactions, hydrogen bonding and van der Waals (enthalpy driven binding). Furthermore, common to BRDs, is a conserved water molecule in the binding site that forms a water bridge between the acetyllysine and an arginine residue. Therefore, displacement of this conserved water molecule in the binding pocket would also contribute to increased binding affinity (entropy driven binding). The library consisted of 3072 unique compounds synthesized by the 'split and mix' method (see chapter 7.2.9 for full materials and methods). There are three points of diversity R1, R2, R3 with 16 *L*-amino acids in position R1, 12 Lysine derivative amino acids in position R2 (lysine derivative structures are given in appendix 8.1.9) and 16 *L*-amino acids in position R3. This leads to 16 x 12 x 16 = 3072 compounds in 16 sub-libraries (see figure 2.18). A sub-library is defined by, after the final amino acid addition, the beads are not mixed, therefore the final amino acid in the sequence is known, and the R1, R2 amino acids are unknown.

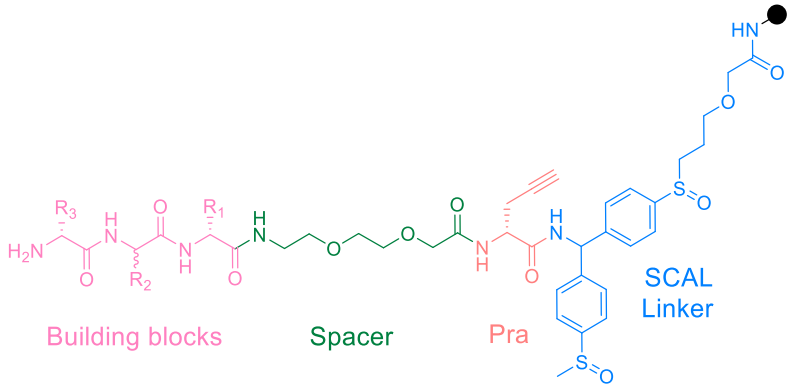


Figure 2.18: Schematic of the OBOC bromodomain focussed library.

Although there are 20 natural *L*- amino acids available for synthesis, four out of 20 are not included in the final library, these are: isoleucine, glycine, cysteine and methionine. The thiol containing amino acids cysteine and methionine are not included as these are highly reactive and can cause side reactions and cross-linking. Glycine tends to be flexible and does not contribute significantly to the binding interactions therefore is not included. And isoleucine cannot be distinguished as a hit from leucine by mass spectrometry as they have identical masses.

The first step in peptide synthesis requires the use of a cleavable linker for subsequent removal of hit peptides for deconvolution. The linker used for on-bead screening has to have orthogonal chemistry for the peptide to remain on-bead yet be capable of removing side chain protecting groups. The linker developed in the Auer lab is a modified version of the SCAL

(Safety Catch Amide Linker), SCAL2, which allows orthogonal deprotection of side chain residues whilst leaving the peptide attached to the resin.[68] After SPPS each sub-library must be analysed for purity and loading capacity. From each sub-library, a small amount of resin (~ 1 mg) was used for quality analysis by fluorescent labelling with tetramethyl rhodamine carboxylic acid (6-TAMRA) on the N-terminus. Ten individual beads were then picked from each sub-library and cleaved from the resin. The cleavage solution was then analysed by HPLC, equipped with a fluorescent detector. Figure 2.19 shows an example HPLC of a single bead cleavage solution, the fluorescent trace is recorded at 555 nm and the purity is analysed by peak area.

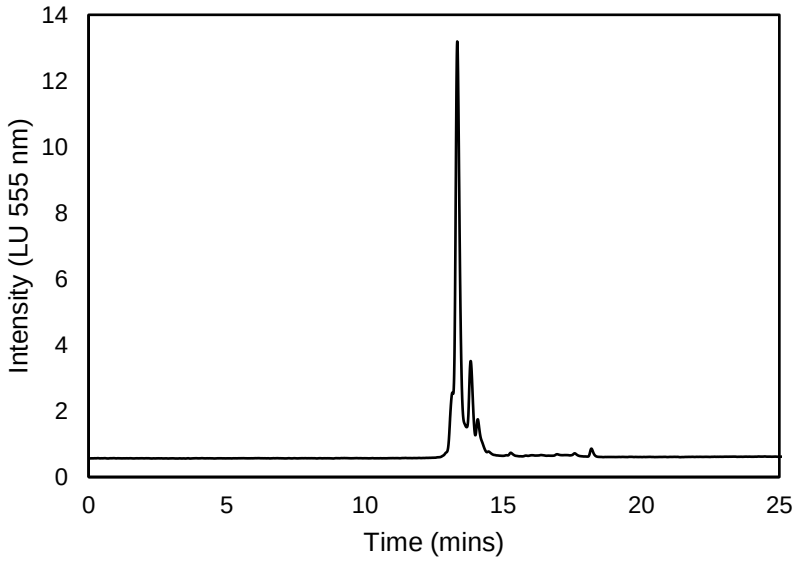


Figure 2.19: HPLC chromatogram of a TMR labelled peptide from a single microbead. The fluorescent trace is measured at 555 nm and purity is measured by peak area (75 % purity).

Figure 2.20 shows a plot of the purity vs the loading capacity of single beads from each sub-library. The information available from the HPLC is two-fold: 1) purity is assessed by the percentage area under the curve and 2) loading capacity is calculated using a standard curve of TMR fluorescence (shown in appendix 8.1.8).

Analysing the purity data, it can be seen that there is an average loading capacity of 17 pmol/bead and an average purity of 60 %. The loading capacity is significantly low, however, a lower compound concentration will inevitably lead to higher affinity peptide binders during screening. The purity is considered acceptable for a screening approach and the impurities may be due to bromination of the peptides under the cleavage conditions. Trimethylsilyl bromide (TMSBr) is used to cleave the peptides from the resin however, it has been shown to

be inefficient and various side products can occur.[120] It is also shown that Trp and Tyr residues are the most problematic for bromination side reactions, as can be seen by the lower purity for these sublibraries (figure 2.20). However, it is generally accepted for QC analysis and for compound deconvolution by mass spectrometry post screening. The final library can also be stored on bead at room temperature in a desiccator for further screening.

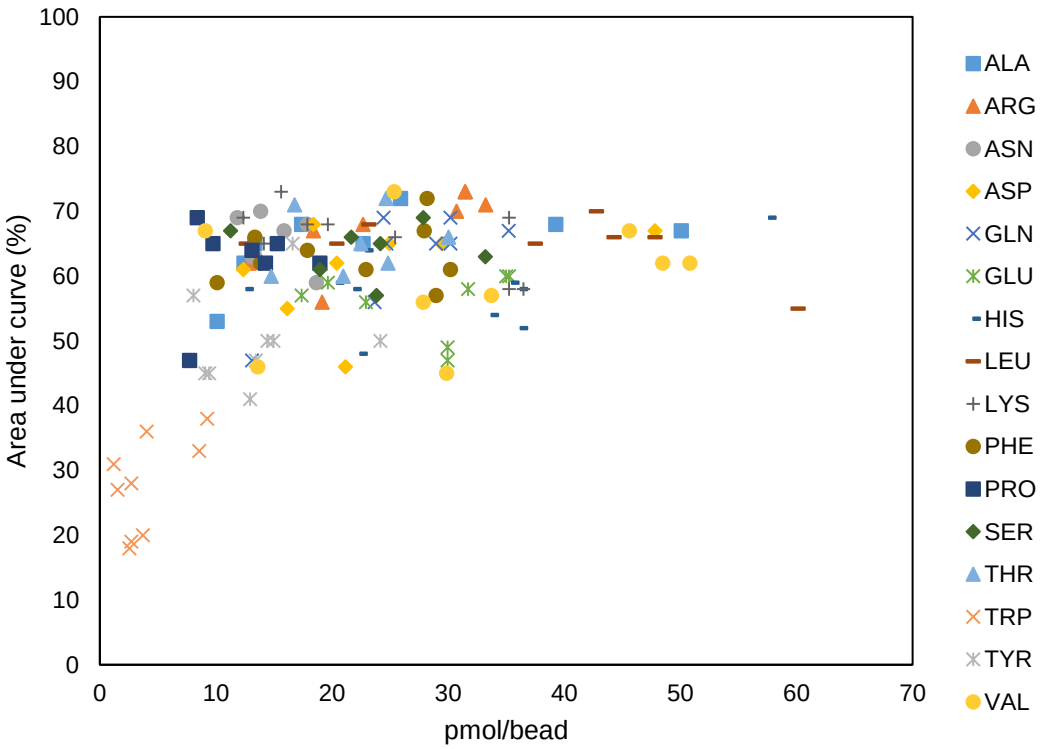


Figure 2.20: Loading capacity of beads in each of the 16 sub libraries plotted vs the purity of the library. Purity is determined by the peak area of the compound by fluorescence detection of the TMR dye (λ_{ex} 555 nm; λ_{am} 575 nm) by HPLC. The loading capacity is also determined using a standard curve of TMR concentrations (see Chapter 7.2.9 and appendix 8.1.8).

2.2.7. Fluorescent Labelling of ATAD2

Maleimide functional groups are often used in the cross-linking of sulfhydryl groups present on cysteine side chains. The cross-linking occurs via conjugate addition into the maleimide which results in a stable thioether linkage. However, any disulfide bonds in the protein must be reduced in order to offer the free thiol for conjugation. Some common, thiol containing reducing agents, e.g. dithiothreitol (DTT), cannot be used in the protein buffer as these will react with the maleimide functionalised dye. Therefore, a different class of reducing agents must be used, such as TCEP (Tris(2-carboxyethyl)phosphine), which was included to prevent any intermolecular disulfide bonds. ATAD2 BRD has two cysteines available on the surface

for labelling with maleimide functionlised dyes. It was chosen to label these cysteines with a cyanine-5 functionalised maleimide, however, attempts to label these sites proved difficult. The labelling reaction is performed in excess of dye (molar ratio 1:7) on a 750 μg scale of protein and overnight at 4 °C.

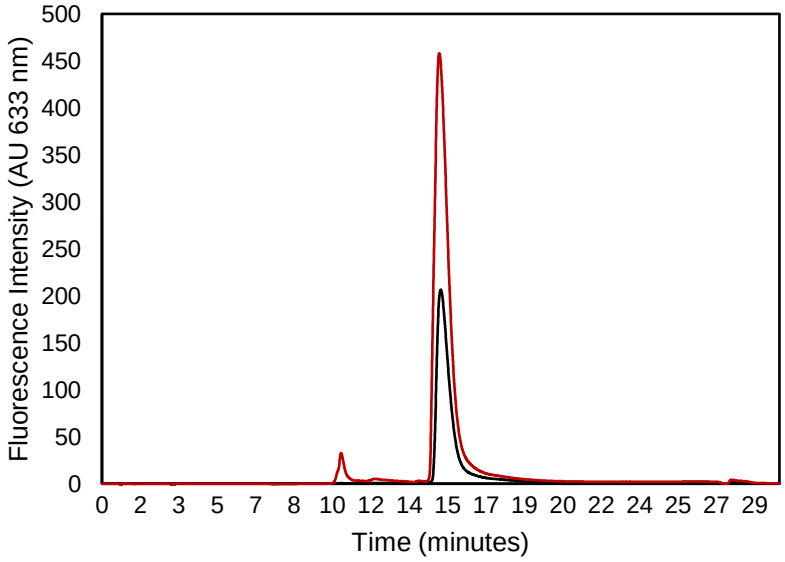


Figure 2.21: HPLC chromatogram overlay of fractions from Cy5 labelling purification. Black line: fraction 2 from the NAP-5 column. Red line: fraction 6 from the NAP-5 column. The protein elutes at ~15 minutes and the free dye elutes at ~10 minutes. The free dye is present in fraction 6, therefore this was not used for pooling. The intensity of absorption is measured at 633 nm.

After the labelling reaction, the protein is purified on a NAP-5 column and fractions are collected by gravity flow in ~50 μl aliquots (see Chapter 7.2.7 for full materials and methods). A NAP-5 exclusion chromatography column is prepacked with Sephadex G-25 for purification of small molecules from large, globular proteins. The free, small molecule dye is retained on the column and the large, labelled protein is separated by gravity flow. The fractions were then analysed by HPLC for presence of labelled protein and free dye. Fractions containing the labelled protein were pooled and concentration was measured by UV/Vis spectroscopy. However, typically, ATAD2 was prone to aggregation in solution, and precipitation at low salt concentrations. Although the labelling reaction appeared to be successful by HPLC analysis (figure 2.21), there was a significant amount of aggregation of the protein as shown by UV/Vis post labelling. This could be seen by stray light corrected concentration determination by UV/Vis analysis (see appendix 8.1.2). Absorption of proteins in the 230 – 300 nm range is dominated by the aromatic side chains, tryptophan (Trp), tyrosine

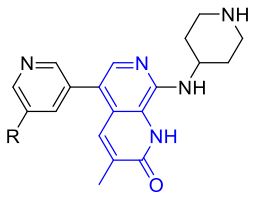
(Tyr) and phenylalanine (Phe).[121] It was not possible to determine the total amount of protein concentration as there is a large absorbance in the 330 nm range. Absorbance in this region is due to light scattering caused by protein aggregation. The calculated pI of ATAD2 protein (151 amino acids) is 5.6 (ExPASY ProtParam), and the labelling reaction is most efficient between pH 7.0 – 7.5, so pH is unlikely to be an issue.

Due to the difficulty in handling the fluorescently labelled ATAD2, it was not possible to screen the OBOC library using this method. Another popular labelling method is to use lysine residues and an N-hydroxysuccinimide ester (NHS) functionalised dye. However, labelling of the protein was most likely corrupted due to protein aggregation in solution therefore this would not likely resolve the issue. An alternative to fluorescent labelling, and detection of binding after target incubation with the OBOC library, would be to use an antibody specific for ATAD2. However, no antibodies exist directly for the bromodomain of ATAD2, therefore this approach is not feasible. However, this particular construct contains a hexa-his tag for purification, and there are a multitude of fluorescently labelled antibodies available specific for his-tag detection. Therefore, this approach would provide a means of detecting hit beads after incubation with the target protein. However, this approach is not pursued as part of this project.

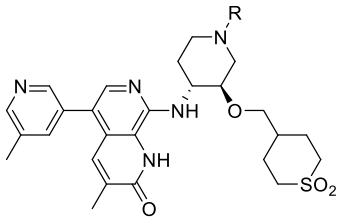
2.2.8. Discovery of small molecule ATAD2 bromodomain inhibitors

Epigenetics is a popular topic and fragments have played a major role in finding lead compounds against several targets, particularly in the BET family of BRDs.[122] At the start of this work there were no known inhibitors or ligands to the particular bromodomain of ATAD2 reported in patent or academic literature. A paper from Fesik *et al* at Vanderbilt University, in 2014 describes the use of SOFAST-HMQC (similar to ^{15}N-HSQC) to screen a huge library of 13800 compounds as mixtures of 12 (1150 individual experiments).[123] This resulted in 65 active compounds with K_Ds ranging from 350 μM to > 2 mM (determined by HSQC titrations), and 12 had affinities below 1 mM. This is a hit rate of < 0.1 % which is low for a fragment screen, although it would be in the normal hit range for a standard HTS. However, several novel chemotypes against this bromodomain were identified. This finding opened up the possibility that the ATAD2 bromodomain is indeed druggable, and almost simultaneously the SGC also published small molecule ligands, albeit with low affinities, for ATAD2.[124] This prompted a surge of interest and in May 2015, two papers from GlaxoSmithKline offered potent inhibitors of ATAD2. The first paper introduced the target-to-lead effort based on the quinolone

fragment hit described in the Fesik paper. Using a time-resolved fluorescence resonance energy transfer (TR-FRET) labelled peptide displacement assay the GSK group screened various analogues of the quinolone fragment against the ATAD2 bromodomain.[85] Then, through various rounds of medicinal chemistry, GSK optimised the series by synthesis of increasingly potent analogues. They also screened the best compounds against the BET family of bromodomains to determine the selectivity profiles. (**2.4,** table 2.2).



2.4



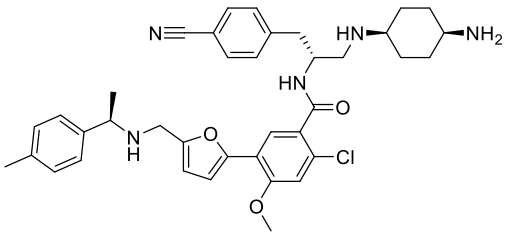
2.5

	R = Me	**R = OMe**	**R = H**	**R = Me**
ATAD2 peptide FRET pIC$_{50}$/LE	5.9(4)/0.31	5.9(4)/0.30	6.9	6.5
ATAD2 SPR K$_d$/LE	2.5 μM/0.30	N/A	7.1 μM	6.7 μM
BRD2 BD1/BD2 pIC$_{50}$	5.2/4.7	5.4/4.7	4.5/< 4.5	<4.3/< 4.3
BRD3 BD1/BD2 pIC$_{50}$	5.1/4.7	5.2/4.7	<4.3/< 4.3	<4.3/< 4.3
BRD4 BD1/BD2 pIC$_{50}$	5.6/4.8	5.9/4.8	4.8/5.3	4.1/< 3.3
BRDT BD1/BD2 pIC$_{50}$	5.1/4.5	5.2/4.6	4.5/< 4.3	<4.3/< 4.3

Table 2.2: Properties of optimised ATAD2 inhibitor compounds. Chemical structures of most potent chemical analogue, the initial fragment is shown in blue. This shows the pIC$_{50}$ values (μM) as calculated using the TR-FRET assay and the comparison to other BET family bromodomains. Reproduced from Demont and Bamborough respectively *et al*.[85,125]

A follow-up paper built on this work used structure-based optimisation (X-ray crystallography) methods to develop this compound further into low nanomolar inhibitors with much greater selectivity profiles (**2.5**, table 2.2,).[126] More recently, in 2017, the SGC reported on the development of a small molecule inhibitor based on a different chemical scaffold to the GSK inhibitors.[127] Purporting to their remit this compound was developed

purely as a tool compound to study the effects of inhibition of ATAD2, and is available for purchase (**2.6**, figure 2.22).



2.6

Figure2.22: Chemical structure of BAY-850 (**2.6**), developed at the SGC in conjunction with Bayer AG Pharmaceuticals.

There are now 37 structures of ATAD2 bromodomain in complex with various ligands available in the PDB. The advent of these small molecule inhibitors of the ATAD2 bromodomain, no observed binding with synthesised peptides and the difficulties encountered in handling the protein led to switching the target focus of this thesis to a separate class of PPI, the ubiquitin conjugating enzyme, UbE2L3. Targeting the family of ubiquitin conjugating enzymes (E2s) is an exciting area of research and, in particular, E2s are a class of enzymes the Auer group has experience in working with.

2.2.9. Conclusions from ATAD2 Project

In summary, ATAD2 was expressed and purified in good yield and purity both in unlabelled and ^{15}N isotopic enrichment labelled form. Two known peptide ligands were synthesised by SPPS and used in ^{19}F and ^{15}N-HSQC NMR experiments, however no ligand binding was observed by chemical shift perturbation. It is possible the ^{19}F probe nucleus was positioned too far away from the binding site and therefore did not interact with the protein, leading to failure of detection of binding. Screening by ^{19}F NMR has advantages as lower protein concentrations can be used, reducing the propensity for precipitation. Also, the ATAD2 ^{15}N HSQC spectrum has not been assigned, therefore any residue shifts in the amide region are unidentifiable. The ^{15}N signal loss on addition of high concentrations of peptide was also an issue, and the high salt concentrations required for protein solubility reduced the benefit of the CryoProbe technology. It was also considered difficult to fluorescently label the protein due to aggregation during the labelling reaction.

The OBOC library was synthesised in reasonable purity and remains available for screening. The objective behind the peptide library design was to deliver biologically active compounds against ATAD2. There are two main categories of screening libraries: random libraries and target focussed libraries. Random libraries are a collection of compounds that have a unique structure and no similarity between known biological actives. Targeted libraries, however, have a distinct similarity between compounds and known biologically active structures, i.e. contain a structural feature recognised by the molecular target, e.g. kinase focussed libraries. They may also contain a privileged scaffold, i.e. a structural motif that interacts with a molecular target. Consequently, this OBOC library with a central lysine derivative, can be thought of as a target focussed library. Essentially, all BRDs bind an acetylated lysine on histone tails, and specificity is only differentiated by the ZA and BC loop regions. Therefore, this library could be screened against any BRD for discovery of novel ligand scaffolds.

2.3. On-protein peptide growing (OPPG)

To continue with the peptide approach to discover novel ligands, a strategy to detect weak affinity binders using ^{19}F NMR was devised. This would entail screening a library of synthesised dipeptides, each containing a fluorinated amino acid. The fluorinated amino acid is used as a sensor molecule to detect weakly binding dipeptides and the most active, i.e. the dipeptide indicating the biggest change in the fluorine resonance, is selected from this library. It was envisaged that a hit peptide could then be extended into a tripeptide by SPPS to increase the affinity (see figure 2.23). A fluorinated tripeptide library can then be screened analogously to identify higher affinity ligands, and so forth. To assess the feasibility of this approach and the potential for hit dipeptide detection, a known peptide ligand was synthesised. The chosen peptide was a modified version of a known tetrapeptide to the cancer-associated, inhibitor of apoptosis protein, survivin.

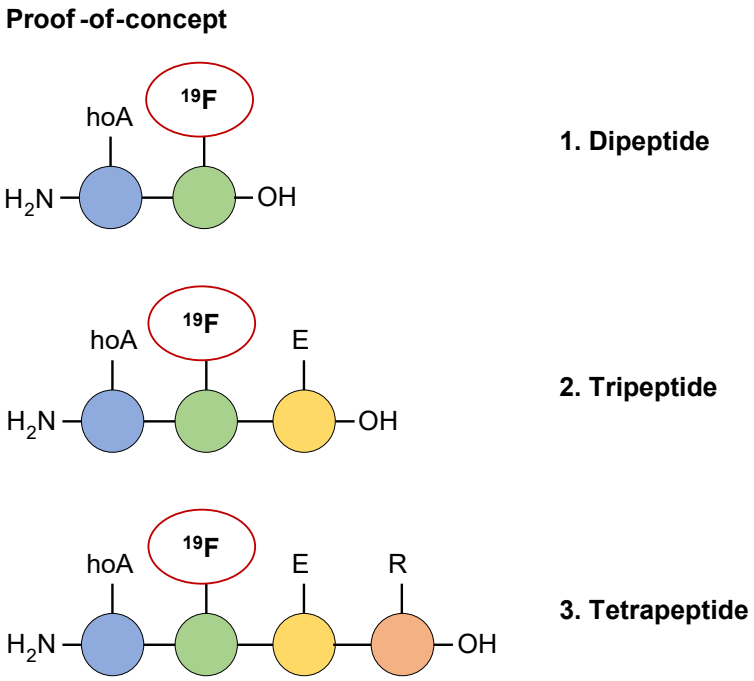


Figure 2.23: Schematic of proof-of-concept peptide growing approach. Coloured circles indicate amino acids, with single amino acid letters. The initial dipeptide (**1**) is screened for binding activity against survivin. The hypothesis is that extension of the initial dipeptide to the tripeptide (**2**) and finally towards the known tetrapeptide (**3**) sequence can be monitored by changes in the 1D ^{19}F NMR assay. The known tetrapeptide sequence is H_2N-hoA-*m*(CF_3)F-E-R-OH. hoA: homoalanine (blue circle); *m*(CF_3)F: *meta*-trifluoromethyl phenylalanine (green circle); E: glutamic acid (yellow circle); R: arginine (orange circle).

2.3.1. Proof of Concept

A collaboration between the Jeyaprakash and Auer groups has identified a non-natural tetrapeptide that binds to the cancer associated, inhibitor of apoptosis (IAP) protein survivin. The structural analysis of survivin complexed with N-terminus of histone H3, identified putative epitopes for other survivin-binding mitotic proteins, such as human Shugoshin 1 (hSgo1). A peptidomimetic of the N-terminal sequence of hSgo1 (native sequence: Ala-Lys-Glu-Arg (A-K-E-R)), was shown to bind to survivin with low micromolar affinity (K_D ~5 μM, PDB ID: 4A0I).[128] This was then modified using the proprietary iterative *in silico* design method, MorPH, (Shave-Auer personal communication) and the original sequence modified to a non-natural amino acid containing tetrapeptide of the form: homoAlanine-carbamoylPhenylalanine-Glutamic acid-Arginine (hoA-cF-E-R; shown in figure 2.24). This peptidomimetic, hoA-cF-E-R, showed 50% increase in survivin binding affinity, compared to

the natural peptide, whilst introducing potential proteolytic stability with a half-life in human plasma of ~ 20 hours.

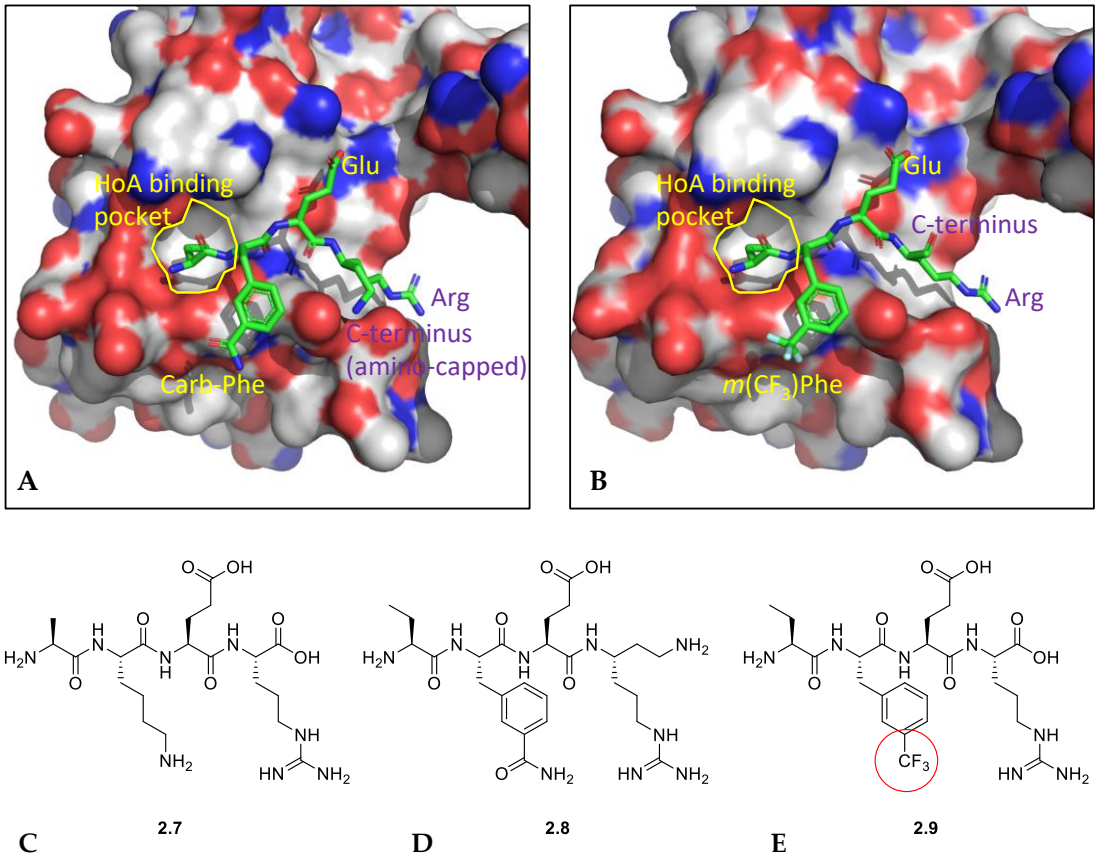


Figure 2.24: Peptide and crystal structures of survivin ligands and complexes. Panel A: X-ray co-crystal structure of peptide **2.8** in complex with survivin. Panel B: predicted binding mode model structure of peptide **2.9** in complex with survivin (Modelling performed by Dr Steven Shave). Panel C: chemical structure of native peptide (**2.7**) (AKER). Panel D: chemical structure of *in silico* designed tetrapeptide (**2.8**) (hoA-cF-E-R). Panel E: chemical structure of fluorinated tetrapeptide analogue (**2.9**) (hoA-*m*(CF_3)F-E-R) proposed to bind to survivin. Single amino acid codes: A – alanine; K – lysine; E – glutamic acid; R – arginine; hoA – homoalanine; cF – carbamoyl phenylalanine; *m*(CF_3)F – *meta*-trifluoromethyl phenylalanine.

The homoalanine variant was proposed to increase affinity by extension of the sidechain into a known binding pocket on the surface. This binding pocket is known to be the main contributor to the peptide affinity. The carbamoyl-phenylalanine (carb-phe) replacement to lysine is a more rigid variant. The original lysine in the peptide contributes no definitive interactions with the protein, as shown by X-ray crystal structures (unpublished), therefore the more rigid carb-phe was proposed to essentially lay across the surface and create a more hydrophobic binding interaction.To utilise this medium-affinity peptide binder and inhibitor of survivin for a proof of concept study for the OPPG approach, designed in this thesis, it was necessary to modify the peptide further with a fluorinated amino acid. It was hypothesised

that introducing a trifluoromethyl group (CF_3) at the *meta* position of the phenyl ring, as a replacement for the primary amide group, in the carbamoyl phenylalanine amino acid would maintain affinity whilst providing a sensitive probe for binding detection by ^{19}F NMR. The position for including a CF_3 probe was based on the crystal structure of the non-natural peptide derivative hoA-cF-E-R (provided by Jeyaprakash *et al*). This shows that the carbamoyl moiety is resting across the surface and does not incur any electrostatic interactions (figure 2.24). The calculated model structure of the fluorinated amino acid replacement is illustrated in figure 2.24 (performed by Dr Steven Shave). This work is as yet unpublished, therefore confidentiality is required. The tetrapeptide, H_2N-hoA-*m*(CF_3)FER-$CONH_2$ (**3.2**, figure 2.29), was synthesised by Fmoc SPPS in good yield and purity (> 95% purity, 11 mg, 85 % yield) (see Chapter 7.2.9 for materials and methods). Figure 2.25 shows the HPLC QC data for the tetrapeptide. See appendix 8.1.7 for mass spec analysis. The C-terminal, primary amide group of peptide 3.2, is a consequence of the cleavage conditions of Fmoc SPPS using the Rink linker. Under these conditions, cleavage with TFA yields a C-terminal primary amide group instead of a C-terminal carboxylic acid, as would be present in a native peptide.

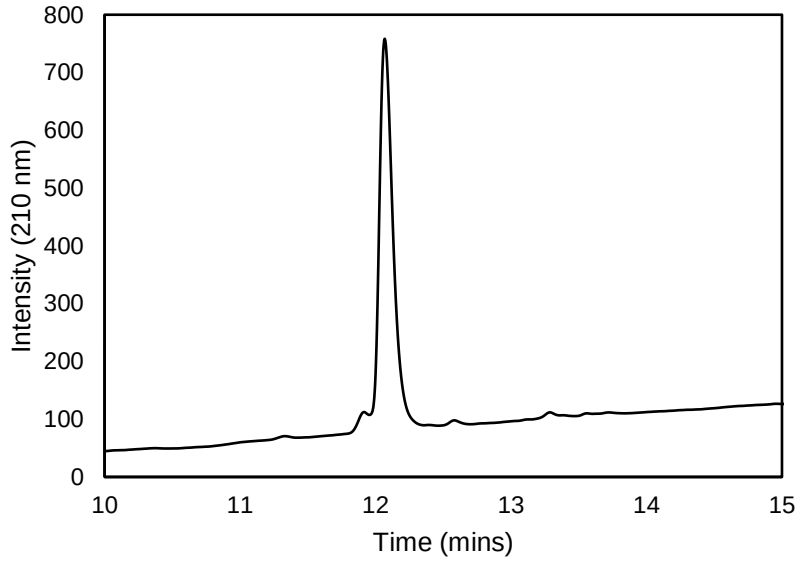


Figure 2.25: HPLC chromatogram of peptide hoA-*m*(CF_3)F-E-R (**3.2**). t_R = 12.1 mins, purity > 95 %.

The binding by ^{19}F NMR was confirmed by spectral analysis of the fluorine resonance in the presence and absence of protein (figure 2.26). There are essentially two observable parameters for binding detection: 1) change in chemical shift of the fluorine resonance and 2) change in

linewidth (coupled to change in intensity). As can be seen in figure 2.26, there is a significant line broadening effect observed in the presence of protein (blue spectrum). The line broadening is caused by a fast relaxation rate of the fluorine nucleus, which occurs due to slow molecular tumbling, this is the case when the ligand is in a bound complex with protein. Assuming a similar K_D of ~ 10 μM of peptide 3.2, to the native peptide 2.8, would mean that ~ 9% of the ligand is complexed with survivin under these experimental conditions. The sensitivity of ^{19}F NMR to ligand binding events is nicely illustrated by the large line broadening effect (figure 2.26). To confirm the binding, and exemplify the sensitivity of ^{19}F NMR to ligand binding events, a titration experiment was performed (figure 2.27). From these data, it can be appreciated that increasing the protein concentration, hence an increase in the fraction of bound ligand, shows an increase in the linewidth of the fluorine resonance (coupled with a decrease in intensity). A detailed description of ^{19}F NMR binding kinetics is given in chapter 4.1.2.

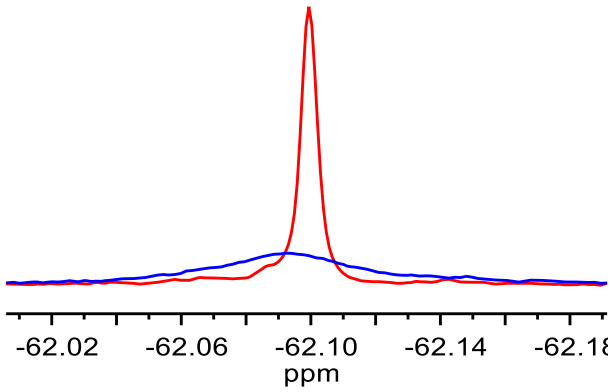


Figure 2.26: A superposition of ^{19}F NMR spectra of peptide 3.2 acquired in the absence (red spectrum) and presence (blue spectrum) of 10 μM survivin. The peptide concentration was 100 μM. Spectra were acquired on a Bruker Avance III 500 MHz NMR spectrometer equipped with Prodigy CryoProbe operating at a spectrometer frequency of 470.7 MHz with a spectral width of 236 ppm and number of scans 256. This yielded experiment time on the order of ~9 minutes.

The tetrapeptide (3.2) was shown to bind by ^{19}F NMR, therefore it was hypothesised that the tri- and dipeptide truncates (see figure 2.29) of the primary sequence might also bind, albeit with decreasing affinity, respectively, and be observable by ^{19}F NMR.

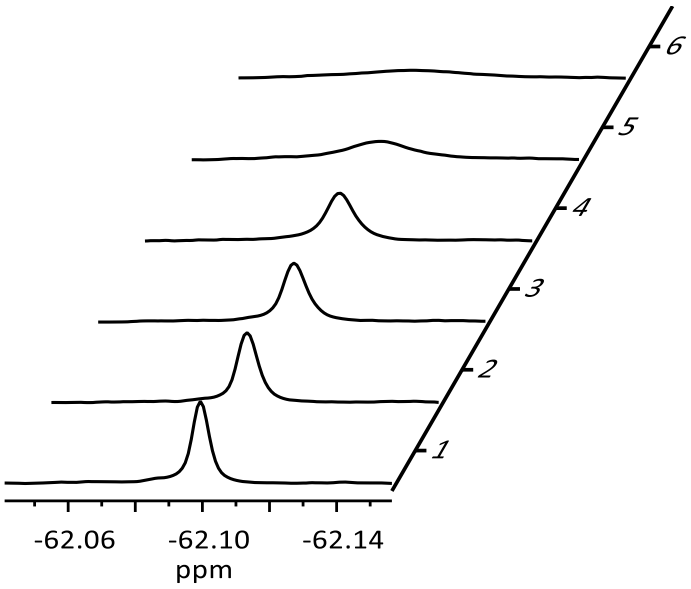
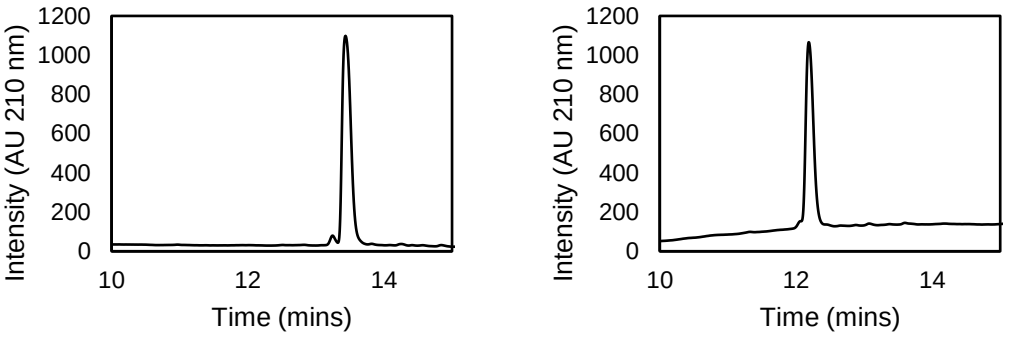


Figure 2.27: Line broadening effect of the fluorine resonance observed on increasing the survivin protein concentration. The peptide concentration is kept constant (100 μM) and protein is titrated in at: 1) 0 nM, 2) 100 nM, 3) 500 nM, 4) 1000 nM, 5) 5000 nM, 6) 10000 nM.

The di- and tripeptide truncated sequences of the fluorinated peptide 3.2 were synthesised by SPPS in good yield and purity (figure 2.28; see Chapter 7.2.9 and appendix 8.1.7).



H_2N-hoA- ***m*****(CF_3)F-OH** (t_R 13.5 mins) **H_2N-hoA-** ***m*****(CF_3)F-E-OH** (t_R 12 .2 mins)

Figure 2.28: HPLC chromatograms of peptides 3.0 and 3.1.

The di- and tripeptide truncates were then screened by 19F NMR against survivin under identical conditions (100 μM peptide concentration and 10 μM protein concentration).

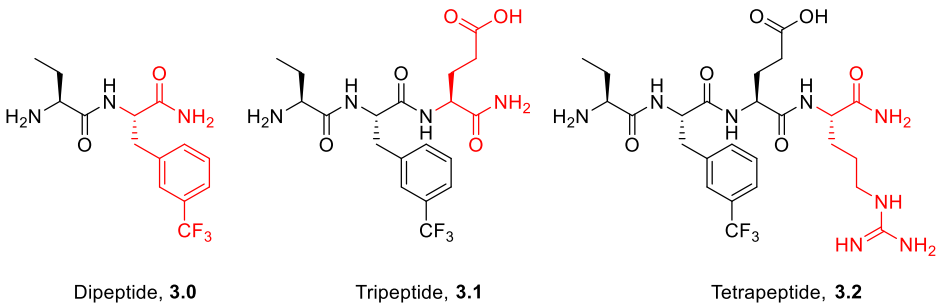
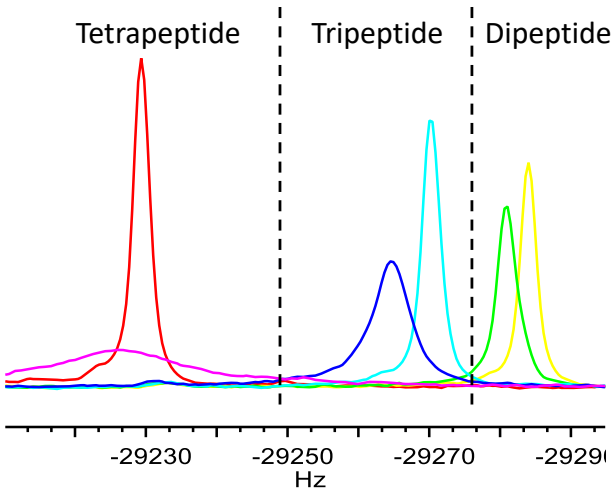


Figure 2.29: Chemical structures of the di-, tri- and tetrapeptide peptidomimetics. Sequences, dipeptide: H_2N-hoA-pfF-$CONH_2$; tripeptide: H_2N-hoA-pfF-E-$CONH_2$; tetrapeptide: H_2N-hoA-pfF-E-R-$CONH_2$. Amino acid codes: hoA – homoalanine; pfF – *para*-fluoro phenylalanine; E – glutamic acid; R – arginine.

The sole ^{19}F resonance for each peptide has a unique chemical shift, and the change in chemical shift and linewidth indicates a binding event which is directly proportional to the binding affinity. Figure 2.30 shows a superposition of the chemical shift spectra for the di-, tri- and tetrapeptides in the presence and absence of protein. It can be seen that the free dipeptide (yellow, figure 2.30), shows a small change in chemical shift ($\Delta\delta$ 2.5 Hz) and reduction in intensity on addition of protein (green). This confirms the binding of the small dipeptide, albeit with low affinity. The free tripeptide (cyan, figure 2.30) shows a larger chemical shift change ($\Delta\delta$ 5.4 Hz) however, also a significant line broadening on addition of protein (dark blue). This confirms that more of the ligand is bound in the complex indicating a higher affinity. The final free tetrapeptide (red) shows extensive line broadening on addition of protein (magenta), almost disappearance of the signal which describes the increase in the binding affinity. Therefore, C-terminal extension of the initial dipeptide results in increased binding affinity which can be monitored by ^{19}F NMR.



Parameter	Dipeptide	Tripeptide	Tetrapeptide
Δδ (chemical shift, Hz)	2.5	5.4	3.4
ΔI_0 (% decrease intensity)	4 %	56 %	81 %
$\Delta LW_{1/2}$ (peak width at half height, Hz)	0.4	2.967	20.8

Figure 2.30: Top: overlay of ^{19}F NMR spectra of the fluorinated truncated peptides in the presence and absence of protein. The free dipeptide in yellow, and with protein in green. The free tripeptide in cyan, and with protein in dark blue. The free tetrapeptide in red, and with protein in magenta. Bottom: Table of changes in observable NMR parameters of the ^{19}F resonance. The change in chemical shift is the difference between before and after the addition of protein.

The aim of this proof-of-concept (PoC) study was to determine that the binding of a fluorinated dipeptide can be observed by ^{19}F NMR and consequently observe increased affinity after successive, iterative extension. Based on this successful "reverse PoC study", it was envisaged that screening a library of fluorinated dipeptides would result in a weak affinity hit which could be extended to a longer peptide with increased affinity.

2.3.2. Application of 'on-protein peptide growing', OPPG, to UbE2L3

The next step was to apply the OPPG method to the highly attractive E2 target UbE2L3 (see section 2.1.8) by generating a library of fluorinated dipeptides and screening against the protein in solution using ^{19}F NMR. This required the synthesis of 16 individual dipeptides of the form shown in figure 2.31. From this point onwards, fluorinated dipeptides are named

from 1 – 16, with the prefix FDL (Fluorine Dipeptide Library), for example, FDL10 (figure 2.32).

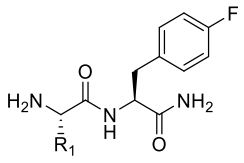


Figure 2.31: Chemical structure of the dipeptide library with a fluorinated phenylalanine and all natural *L* amino acids at position R1.

The fluorinated amino acid chosen for this library was *para*-fluoro-phenylalanine (pfF) due to wide availability and also, to assist in purification. The phenylalanine allows visualisation of the peptide by UV absorption in HPLC analysis and, hence, in determination of purity. Also, a single fluorine nucleus in the *para*-position is not thought to cause significant change in structure or electronic effects therefore it is close to the native form of the amino acid Additionally, although a single fluorine is less intense than a CF_3 group, aryl fluorines (i.e. F atom connected to a ring system as in pfF) are more sensitive than aliphatic fluorines (CF_3) to the change in environment. This gives rise to a larger chemical shift range and, hence, a more sensitive detection method. Solid-phase peptide synthesis traditionally constructs a peptide from the 'C' terminus to the 'N' terminus in a linear fashion. Using a solid support, such as Tentagel™ resin the 16 dipeptides were synthesised in good yield and purity (see chapter 7.2.9 for materials and methods). Each peptide was quality checked by orthogonal, biophysical methods, including HPLC, LCMS and ^{19}F NMR. The structures and QC data for each peptide dipeptide are given in appendix 8.1.10, an example for FDL10 is shown in figure 2.32.

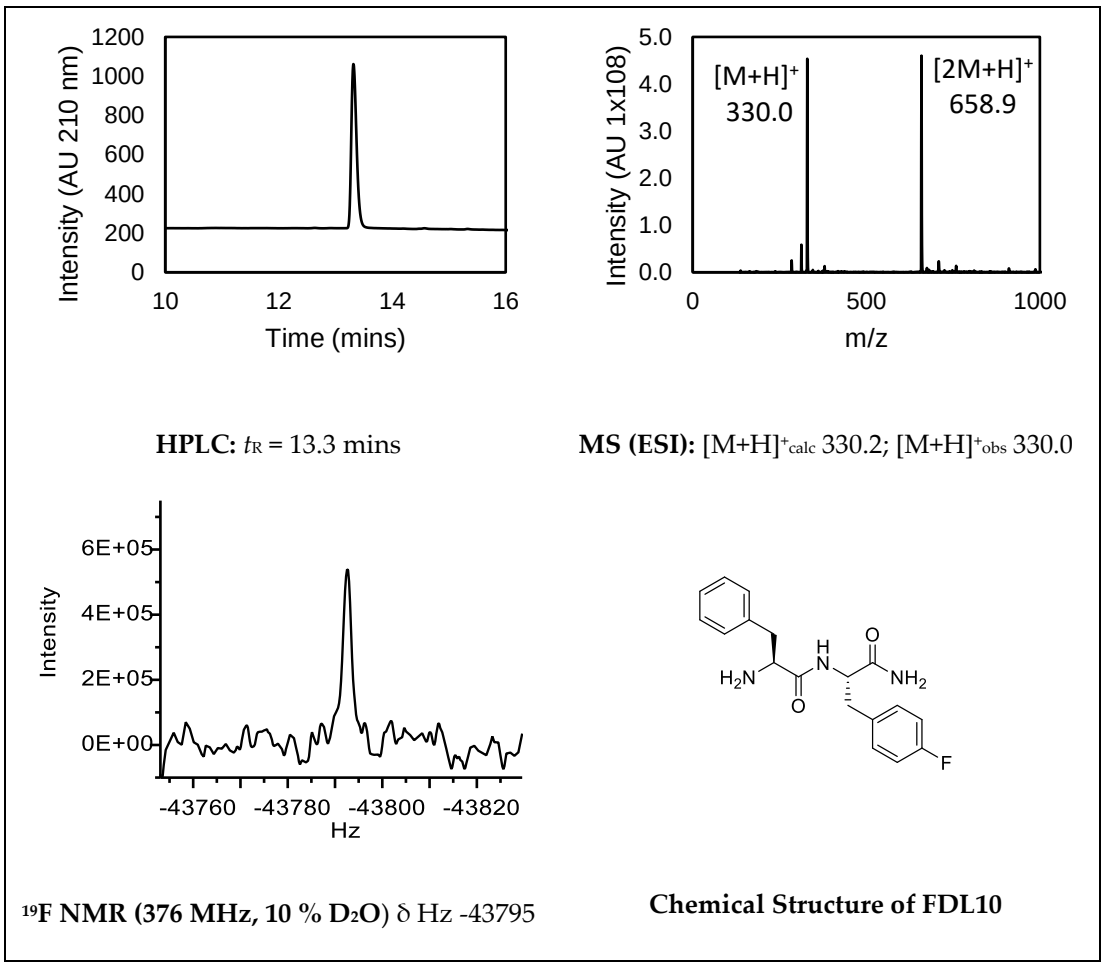


Figure 2.32: Example quality control data (QC) for FDL10 (H_2N-Phe-pfPhe-$CONH_2$).

2.3.3. UbE2L3 expression and purification

The protein UbE2L3 was expressed and purified in good yield by traditional shake flash bacterial expression methods. The yield of protein was 2 $mg.L^{-1}$ from 1 L of bacterial culture and 95 % purity as measured by HPLC. The protein was also analysed by SDS-PAGE for identification by molecular mass. All materials and methods are given in chapter 7.2.3 and QC data is given in appendix 8.1.3.

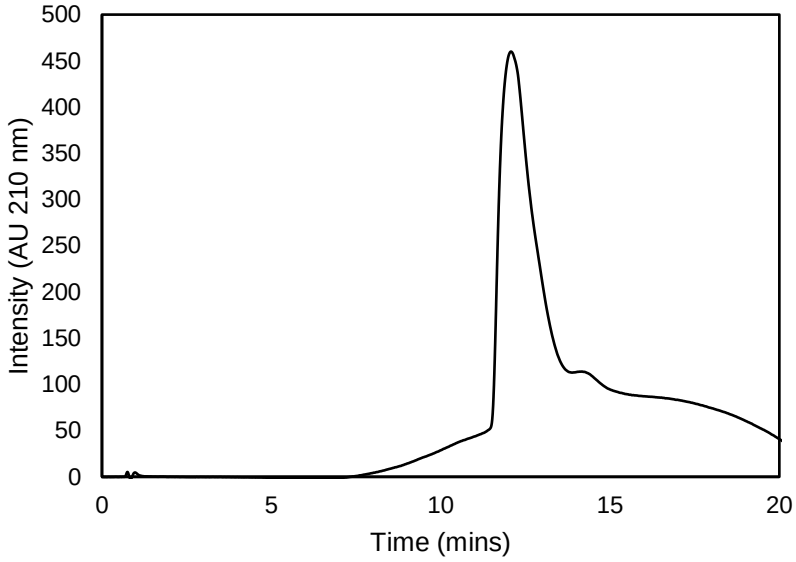


Figure 2.33: HPLC chromatogram of UbE2L3. 10 μg was injected onto the column for analysis and the observed t_R = 12.1 mins and > 95 purity as measured by peak area.

2.3.4. ^{19}F NMR screening of the fluorinated dipeptide library (FDL)

The fluorinated dipeptides were pooled based on minimal overlap of ^{19}F resonances. This was to minimise the amount of protein needed for the screen. At 10 μM protein concentration equates to 5 nmoles required per compound pool (volume 500 μl) which is 100 μg of UbE2L3 (M_w 19.5 kDa). Figure 2.34 shows an example pool of the fluorinated dipeptides in the presence (blue) and absence (red) of protein. Due to the solid-phase cleavage conditions of peptides results in production of the compound as a TFA (trifluoroacetic acid) salt. TFA is observed in the 1D ^{19}F NMR spectrum (figure 2.34, top left) however, this is useful as it can be used as an internal resonance reference shift. That is, in the superposition spectra the TFA peaks are overlayed so that any shift in the peptide resonances can be attributed to binding and not to minute changes in pH. The lower spectrum in figure 2.34 shows the fluorine resonances corresponding to the dipeptides. There are very slight changes in the resonance signal on addition of protein which could indicate a binding event. However, there are changes in each dipeptide which could also indicate nonspecific interactions of each peptide with the protein. Figure 2.35 shows a bar chart plot of the chemical shift difference of each dipeptide on addition of protein. The largest change in chemical shift was deconvoluted to the H_2N-Val-*p*fF-$CONH_2$ dipeptide (FDL16) with Δδ 1.4 Hz. This is a very weak change in chemical shift and is likely a > 1 mM affinity ligand. Somewhat interestingly, the next highest observed chemical shift change is due to FDL08 or the H_2N-Leu-*p*fF-$CONH_2$ dipeptide (Δδ 1.3 Hz, $\Delta LW_{1/2}$ 0.2 Hz).

Leucine and valine are both classed as branched-chain amino acids with aliphatic side chains. This could indicate, a hydrophobic interaction with a small pocket or a non-specific interaction with the surface of the protein.

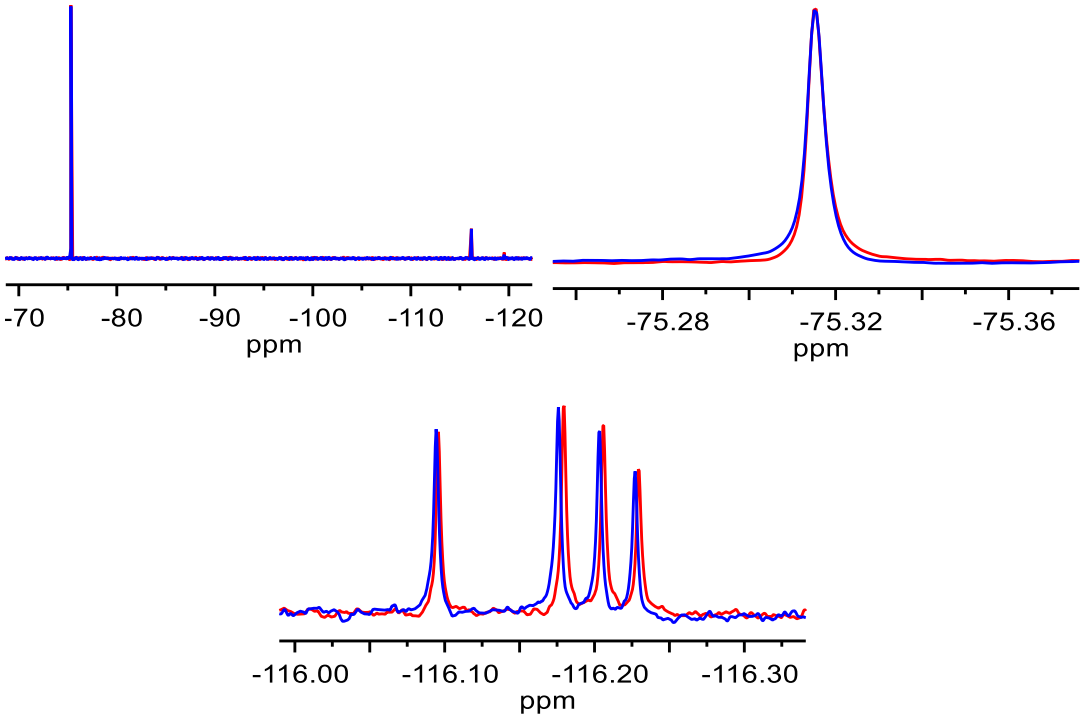


Figure 2.34: ^{19}F NMR screening of the fluorinated dipeptide library vs UbE2L3. Top left: Superposition of compound pool (100 μM each) in absence (red) and presence (blue) of 10 μM UbE2L3. The peak at -75 ppm is TFA used as a reference peak. The cluster of peaks at ~115 ppm are the peptide peaks. Top right: close up of the superposition of the TFA peak in the compound pool. This peak shows precise overlap in signal indicating no binding of TFA to the protein, this is then used as a reference for the compound peaks. Bottom: 4 fluorinated dipeptide compounds (100 μM each) in absence (red) and presence (blue) of 10 μM UbE2L3. There are very small downfield shifts detected in some of the resonance signals.

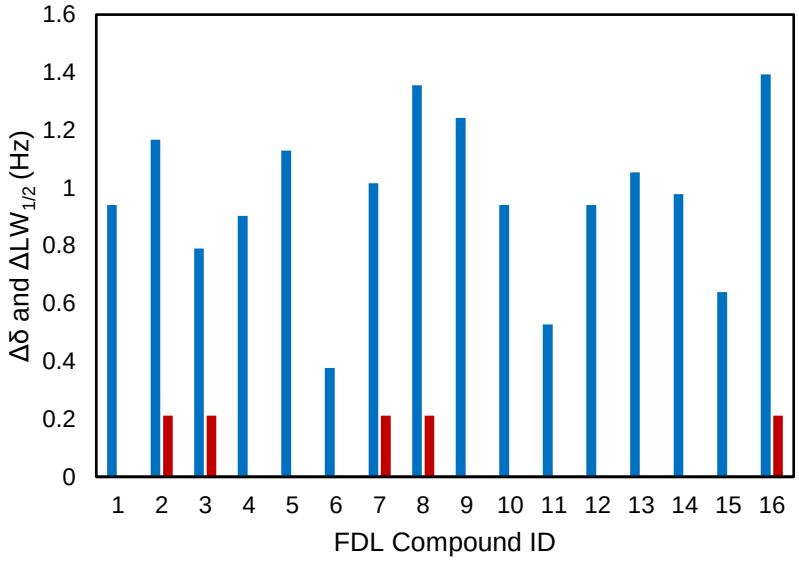


Figure 2.35: A bar chart plot of the ^{19}F resonance shift (blue bars) and corresponding line broadening (red bars) of each fluorinated dipeptide (FDL) in the presence of 10 μM UbE2L3. The shifts are weak indicating a low affinity.

To validate the observed binding effect, the experiment was repeated with FDL16 and independently measured in the presence and absence of protein. Again, in this experiment a similar Δδ 1.5 Hz chemical shift change was revealed (figure 2.36). This shift is comparable to the chemical shift expected, as the proof-of-concept dipeptide (compound **3.0**, figure 2.29), upon binding to survivin, showed a chemical shift change (Δδ) of 2.5 Hz. FDL16 was then taken forward for chemical extension at the C- and N-termini by SPPS.

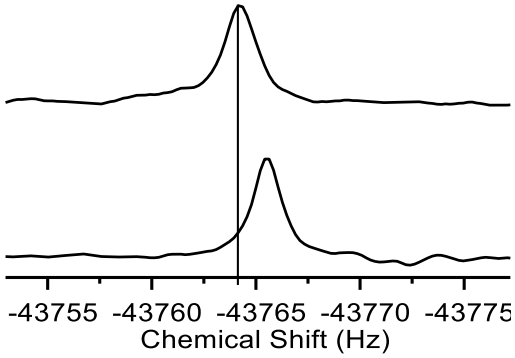


Figure 2.36: Stacked spectra of fluorinated dipeptide H_2N-Val-*p*fF-OH in the absence (bottom) and presence (top) of 10 μM UbE2L3. Peptide concentration was 100 μM. There is a slight change in chemical shift (Δδ 1.4 Hz, $\Delta LW_{1/2}$ 0.28) and decrease in intensity on addition of protein indicating a likely binding event. The dashed line indicates the downfield shift in fluorine signal on addition of protein.

2.3.5. ^{19}F NMR screening of the fluorinated tripeptide libraries

The fluorinated dipeptide, FDL16, was then used as a scaffold for extension in an N and C direction (see figure 2.37). Using Fmoc SPPS a library of 32 fluorinated tripeptides were synthesised in good yield and purity (see chapter 7 for materials and methods). Figure 2.38 shows the chemical structure C-FTL10 and corresponding QC data as an example (see appendix 8.1.11 and 8.1.12 for QC data of all synthesised tripeptides). The tripeptides were then pooled again based on minimal overlap of ^{19}F NMR signals (see appendix 8.1.14 and 8.1.15 for all peptide pools). The peptides are sorted based on their fluorine resonance, the first peptide is placed into pool 1, the second into pool 2 and so forth until the desired number of pools is reached. From herein, the N- and C-terminal extended tripeptides are referred to as N-FTL and C-FTL respectively (i.e. N-terminal Fluorine Tripeptide Library).

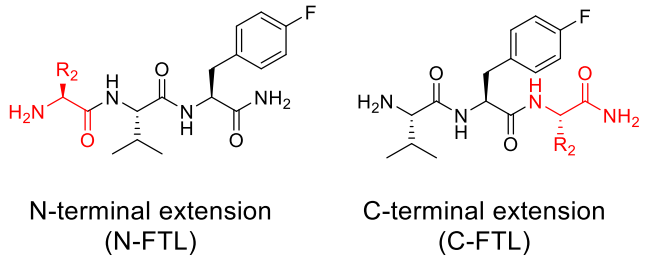


Figure 2.37: Chemical structures of the N- and C-terminally extended tripeptide libraries (N-FTL and C-FTL respectively highlighted in red). R2 is the position of the 16 natural *L* amino acids.

Again, the peptides were screened against UbE2L3 under identical conditions, 100 μM ligand and 10 μM protein, as used in the dipeptide screen, and analysed for chemical shift changes or line broadening effects. However, changes in the fluorine resonances were not as significant as the dipeptide library pools, i.e. extension at either the N or C terminus did not contribute to an increase in binding affinity. The largest change in chemical shift occurred *via* N-terminal extension with a tryptophan residue (compound N-FTL14) with Δδ 0.9 Hz. However, this is a smaller change than the original dipeptide, and there were no line broadening effects observed. Figure 2.39 shows a bar chart illustrating the chemical shift changes for the N-FTL and C-FTL peptide libraries. This leads to the conclusion that both N- and C-terminal extension hinders ligand binding. This could indicate a rather small binding pocket, such that a small trimeric ligand is too large for the binding site. The E2 enzymes are generally classed as difficult, or undruggable targets due to the overall globular structure and absence of a well-defined binding pocket.[129]

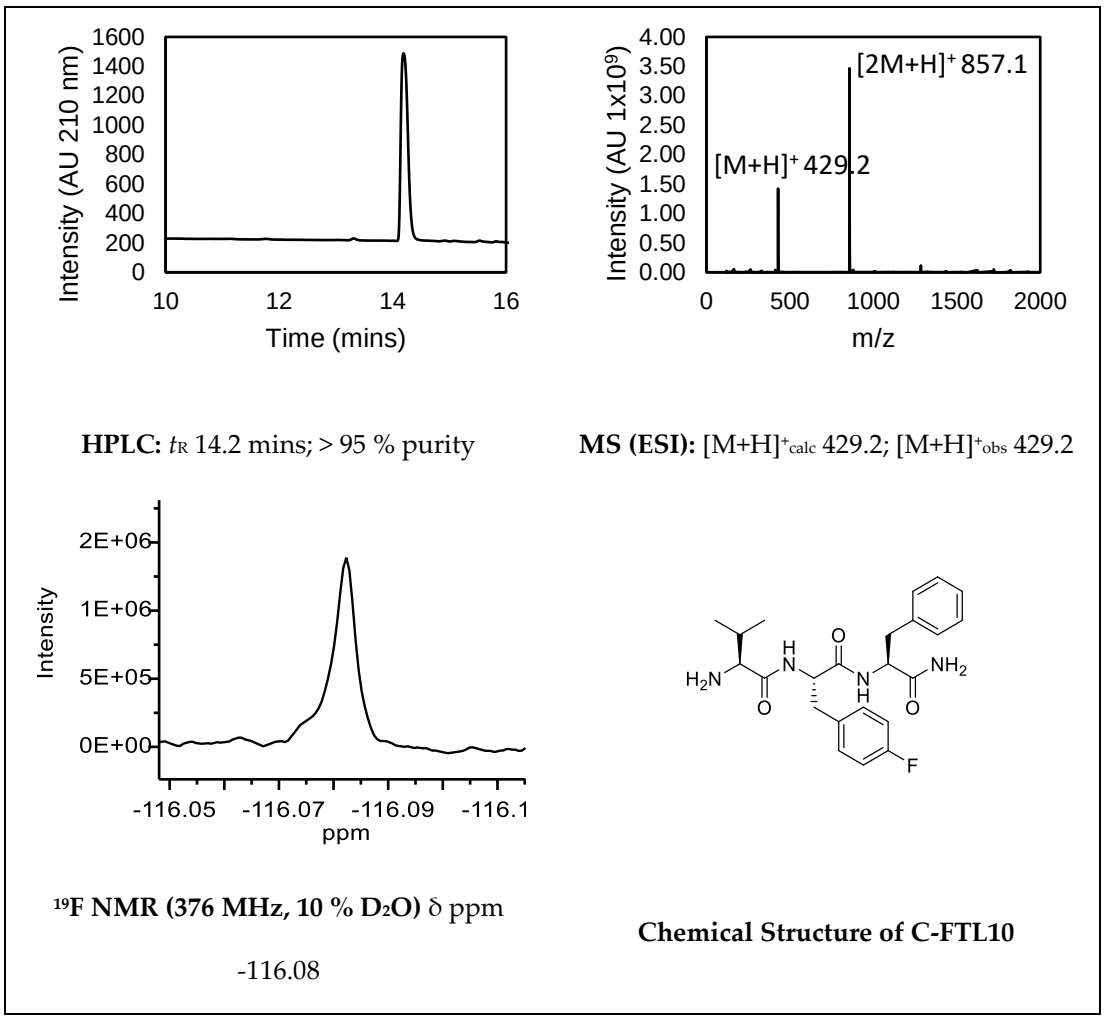


Figure 2.38: Example quality control data (QC) for C-FTL10 (H_2N-Val-pfPhe-Phe-$CONH_2$).

The result indicating that the N-terminal or C-terminal extension of the weakly binding peptidomimetic dipeptide with 16 natural amino acids did not lead to increased, but to decreased binding affinity was surprising. Taking the best of the binding tripeptides and replacing the natural amino acids with non-natural amino acids appears to be a possible step forward. Also, the effect of capping, and replacement of amino acid side chains with bioisosteres is a further option. Although of low affinity, it might be useful to test the best identified binders in a ^{15}N-HSQC experiment to identify a binding site on UbE2L3. This would then allow docking experiments to better understand SAR situations and plan further chemistry.

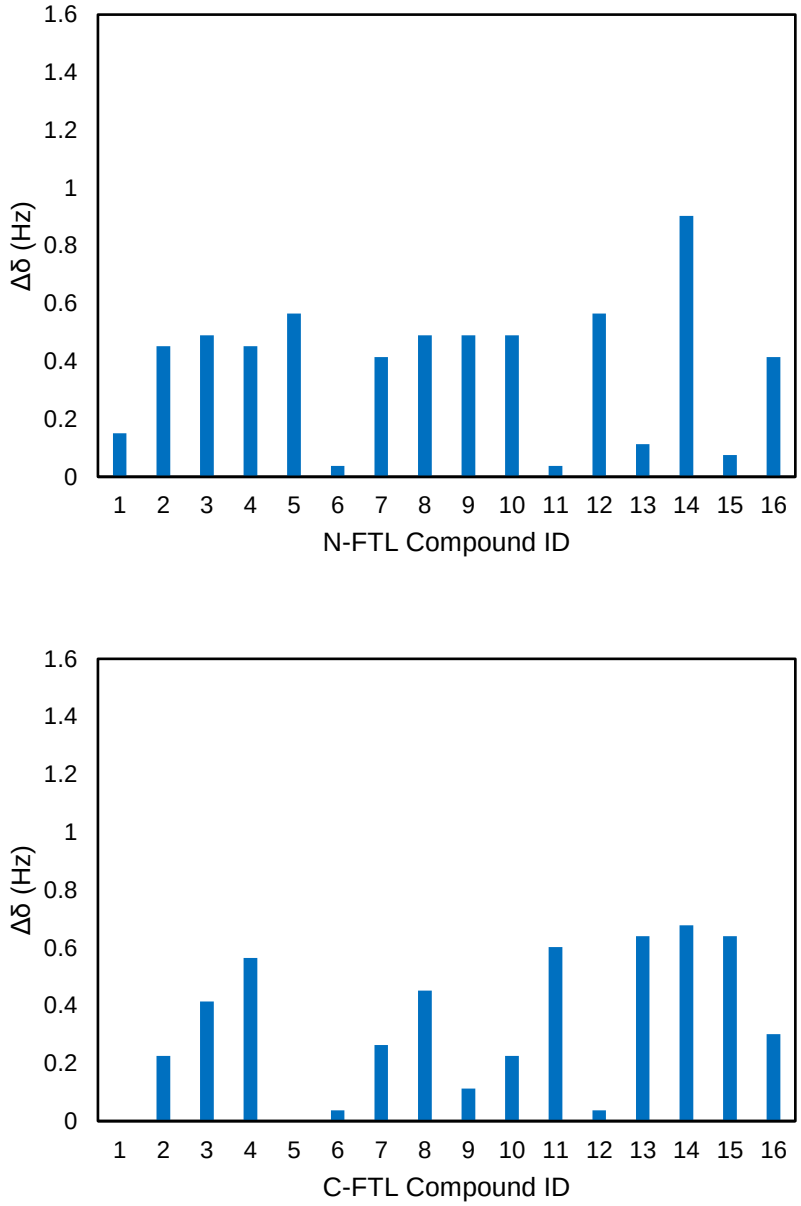


Figure 2.39: Chemical shift change for fluorinated tripeptide libraries. Top: A plot of the resonance shift of each fluorinated tripeptide (present at 100 μM) from the N-terminal extension library (N-FTL) in the presence of 10 μM UbE2L3. Bottom: A plot of the resonance shift of each fluorinated tripeptide (present at 100 μM) from the C-terminal extension library (C-FTL) in the presence of 10 μM UbE2L3. The shifts are very weak (< 1 Hz) indicating a very low affinity.

There are a number of small molecules targeting E2 enzymes which classes this family of enzymes as druggable (see chapter 4.1.3.1). There is also one compound that has a broad spectrum activity on E2 conjugating enzymes. The small molecule compound BAY 11-7082 inhibits the formation of E2-ubiquitin conjugates by direct covalent interaction in the binding site (see figure 2.40).[130] However, recent work has uncovered the broad spectrum activity of

this compound on multiple members, not limited to E2 enzymes, of the UPS. Specifically, it has been found to have target activity against deubiquitinases and multiple other proteins.[131]

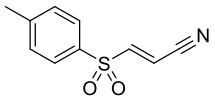


Figure 2.40: Chemical structure of compound BAY 11-7082. This covalently attaches to the active site cysteine in E2 enzymes.

The average molecular mass of a dipeptide is ~ 300 Da, which is also 'fragment' sized (described in chapter 3 onwards). Therefore, in the future the OPPG technique will be further explored in the context of ^{19}F fragment based screening (FBS). The focus of this thesis from herein switches to fragments and screening of low molecular weight compounds for ligand discovery. The UbE2L3 protein is the target focus for fragment screening presented in the following chapters.

2.4. Conclusions and future work

A successful proof-of-concept study was achieved with detection of a weak affinity dipeptide peptidomimetic to survivin by ^{19}F NMR. It was also shown that increased affinity towards survivin could be observed using ^{19}F NMR by extension of the dipeptide to the corresponding tripeptide and tetrapeptide. The medium affinity tetrapeptide is a very useful probe molecule for future studies, and the use of ^{19}F NMR is a simple yet sensitive detection method. For future work, this fluorinated peptide could be used in a FAXS approach (described in chapter 4.1.1) for competition binding experiments in a fragment-based screen.

The aim was to use the knowledge acquired via the proof-of-concept peptidomimetic study on survivin to develop a fluorinated dipeptide library using the UbE2L3 target in an initial approach. Using solid-phase peptide synthesis (SPPS) a 16 member fluorinated dipeptide library was synthesised for ^{19}F ligand-observed NMR screening. A fluorinated amino acid, a *para*-fluoro-*L*-phenylalanine (pfF), was used as the fluorine containing moiety and these were screened against the UbE2L3 protein. However, initial hit dipeptides showed very low binding affinity and maybe binding unspecifically. The most active hit dipeptide was taken forward and a library of 32 C- and N-terminal extended tripeptides were synthesised by SPPS. These were screened for activity against UbE2L3 by ^{19}F NMR however, binding was even

weaker than the dipeptide. This confirms that UbE2L3 protein is a difficult target for ligand discovery.

To develop the OPPG approach further, other E2 enzymes could be screened with the initial fluorine dipeptide library. This would provide a measure of the druggability of this family of proteins and offer vital information for the pharma industry. Also, this approach can be extended to using different fluorinated amino acids as the chemical probe and non-natural amino acids in the 'N' and 'C' positions. For example, there are multiple fluorinated isomers of tryptophan that are commercially available which could be used as the fluorinated amino acid probe. Again, this is scope for creating a PPI target focussed library, for example, Bogan and Thorn (1998) compiled a database of 2325 alanine mutant proteins, for which the change in free energy of binding in the PPI has been measured.[132] They describe the free energy of binding is not evenly distributed across the PPI interface, and binding is made up of hotspots. The amino acid residue most likely to occur in a hotspot is tryptophan,[133] therefore using this amino acid as the probe molecule might increase the chances of finding a novel ligand.

This novel approach to 'on-protein peptide growing' OPPG, can theoretically be used against any target. The simplicity of peptide synthesis, coupled with the assay sensitivity and ease of set up could be an attractive option to other academic groups looking to explore new peptidic ligands.

Chapter 3

Fragment Library Design

This chapter describes the approaches used to generate fragment libraries for use in NMR-based screening. An introduction to library design principles and strategies is given along with some examples from Pharmaceutical and academic groups. A short analysis of the properties of available commercial libraries is also presented. Following this, three strategies which were successfully applied to design and generate fragment libraries for purchasing unique compound collections are described.

I would like to acknowledge Dr Steven Shave for the huge contribution to the cheminformatics work described here. Where stated the programming, indexing and analysing of computational data is completed by Dr Steven Shave. I would like to thank him for his generous help and insight to produce these fragment libraries which provided an essential element for successful data generation in this thesis work.

3.1. Introduction

The fundamental tenet behind fragment-based screening is that small fragments can more effectively probe chemical space.[134] A range of biophysical methods can be used to screen for hits, as described previously (Chapter 1.3), and, as a consequence, the library must be designed in such a way to limit certain physiochemical properties of fragments whilst maintaining a maximal level of diversity.[25] According to the Hann model of molecular complexity, there is an increased chance of detecting binding when screening small, low complex fragments with a high sensitivity than large compounds with low sensitivity.[135] As molecules become larger and more complex, they are more likely to have a single binding mode. This is shown by Schuffenhauer, in an analysis of the Novartis HTS data, where an increased hit rate is observed for screening low-complex fragments with higher sensitivity.[136] Fragment library design is extensively covered in the literature[134,137] and the most important principles are discussed in section 3.1.1.

3.1.1. Library Design Principles

The success of a FBDD campaign is greatly dependent on the design of the library, i.e. the quality of the library dictates the quality of the hits found.[138] The fragment quality, for example, can be defined by the capacity for chemical follow up into lead compounds. Another important consideration is the specific technology to be used for screening, that is, the design must ensure fragments are amenable to the type of assay that will be used. For example, many assays prove unsuitable for screening, with false positives a result of compound aggregation at the high fragment concentrations required. Additional considerations that must be taken into account in the overall library design include: the desired physicochemical properties of fragments to be included, the aqueous solubility and purity, an assessment of molecular diversity, the chemical tractability of the fragments for follow-up, which chemical functionalities are disallowed, the drug-likeness of the fragments and sampling of privileged medicinal chemistry scaffolds. Another aspect to library design is the debate over the optimal size, number and shape of fragments to be screened which, in particular, arouses controversy.[139] Although all these factors must be taken into account when designing a fragment library, there are three main aspects to qualitative library design that are universally accepted, these are: chemical space coverage, chemical tractability and few reactive moieties. Each of these parameters are discussed below.

3.1.2. Chemical Space and Rule of Three

Chemical space is a concept that encompasses all possible molecules and chemical compounds that adhere to physical construction principles. Theoretically, there are 10^{60} – 10^{100} small molecule compounds available that populate this huge chemical space,[140,141] and it is therefore impossible to methodically search for the optimal molecule. A fundamental advantage of FBS is that a better exploration of chemical space is available with smaller ligands as a result of lower molecular complexity and lower inherent selectivity. Fragment libraries are therefore commonly designed to maximise chemical diversity, whilst limiting certain physical properties of compounds. Ultimately, many of the library design strategies use similar property filters to ensure the quality of fragments and to maximise diversity. Lipinski's Rule of Five, were the original guidelines for the development of a bioavailable drug compound. These parameters have been developed over time to include rotational bonds and polar surface area. With the advent of fragment-based screening, researchers at Astex observed that hits identified in screening campaigns appeared to obey a 'Rule of Three' (Ro3). The Ro3 states that a fragment has:

- < 300 Da molecular mass
- ≤ 3 cLogP (partition coefficient)
- ≤ 3 hydrogen bond acceptors
- ≤ 3 hydrogen bond donors
- ≤ 3 rotational bonds (NROT)
- ≤ 60 $Å^2$ polar surface area (PSA)

These parameters are widely considered as rough guidelines,[142] however, academia and pharmaceutical companies tend to follow them in design of their in house libraries. Often the number of hydrogen bond donor and acceptor pairs are higher. However, it is shown that a minimum molecular mass of 150 Da is required to minimise the effect of reorientation of the fragment during hit development.[143] The Ro3 has had a major impact in drug discovery however, recently has come in for some criticism. Köster *et al* designed a fragment library of 364 compounds without strict conformation to the Ro3.[144] They screened the library against endothiapepsin and eleven crystal structures were determined covering fragments with diverse binding modes. They noted that only 4 of the 11 fragments are consistent with the Ro3, therefore restrictions to this rule are limiting the pharmacophoric variety of chemical hits. In general, the average molecular weight of libraries is falling however, it is the physiochemical

properties, e.g. solubility and stability, and the potential for development that drive library design.

3.1.3. Chemical Tractability

Fragment libraries require additional constraints in design methods owing to the high aqueous concentrations needed for biophysical binding detection and the need for chemical tractability of hit compounds. Chemical tractability describes the capacity of fragment-like compounds to be developed into drug-like molecules. This is an important consideration in the design of fragment libraries as rapid development of hit compounds is paramount in a competitive field. One challenge in FBS is that polar moieties (such as amines or acids) in low molecular mass fragments often provide the key interaction, therefore any synthetic change to this group (for example amide formation) could eliminate the binding interaction.

3.1.4. Reactive, Toxic and PAINs Compounds

Compounds containing specific chemical functionalities may react covalently with biological targets. There are examples of this in certain drugs, however, usually these molecules are excluded from screening libraries. In a review by Rishton, the chemical groups illustrated in Figure 3.1 were reported to be potentially reactive.[145] Although the functionalities in Figure 3.1 are generally avoided, some chemical groups have been found to be frequent hitters in high throughput screens. In 2010, Baell and Holloway introduced the term Pan-Assay Interference Compounds (PAINs) to describe a number of substructural features that appear to hit multiple targets, these are illustrated in Figure 3.2.

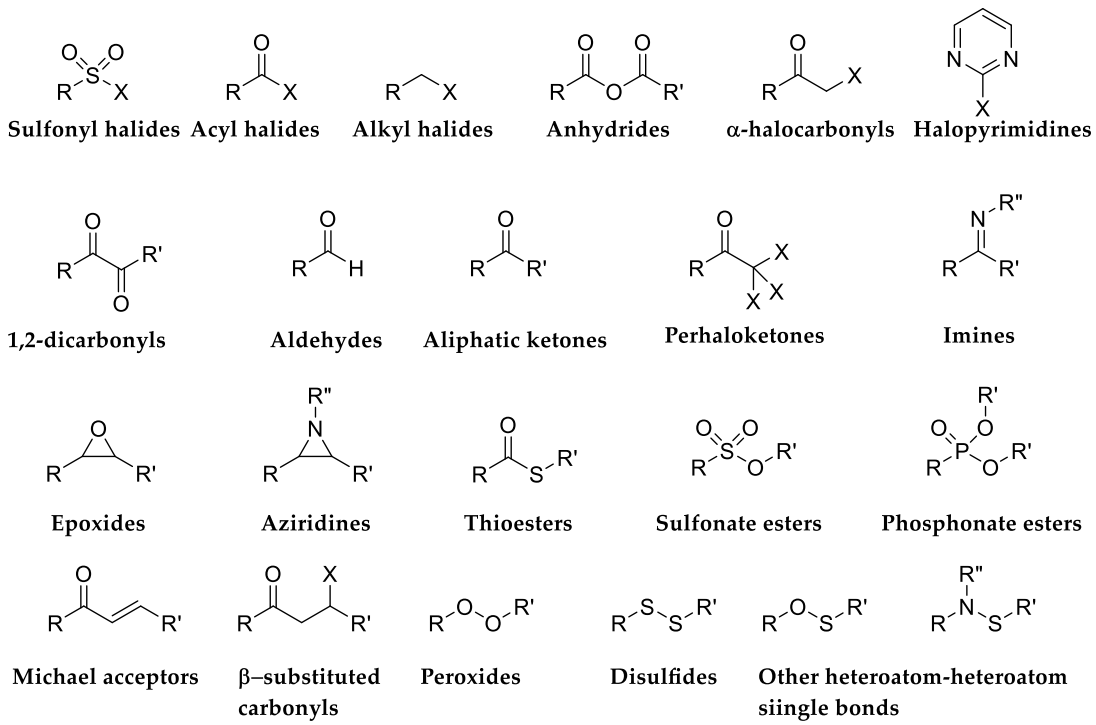


Figure 3.1: Chemical structure of potentially reactive chemical moieties. These functionalities are potentially reactive and are usually avoided in fragment libraries. Reproduced from Rishton.[145]

A common feature among many PAINs moieties is that they contain a Michael acceptor, which can react with nucleophilic residues in proteins, for example, rhodanines. Rhodanines are the archetypal example of the nature of the problem. A search of the literature reveals 2,132 rhodanines that reportedly have biological activity in 410 papers, from 290 organisations, of which 24 are commercial companies.[146] It is shown that these types of compound can undergo light-induced reactions that irreversibly modify proteins. They can bind either reversibly or irreversibly, however, the tendency to form covalent adducts make them unsuitable for drugs or probe molecules. They have since been reported to be active in other assays and are frequently missed in reactivity filters. Therefore, since its inception, PAINs filters have become a common implementation in fragment library design. It is important to note, however, that the structural features in PAINs compounds were identified using only one biochemical assay. The inclusion of PAINs filters are facing increasing criticism with over 80 small molecule FDA-approved drugs contain PAIN structures.[147]

The design principles described above are generally adopted by academic groups and the Pharma industry for fragment-based screening campaigns. Applying these considerations in the final library design offers the highest chance of identifying fragment hits. A summary of

library design approaches adopted by various academic and pharmaceutical companies is given in the following section (3.1.5).

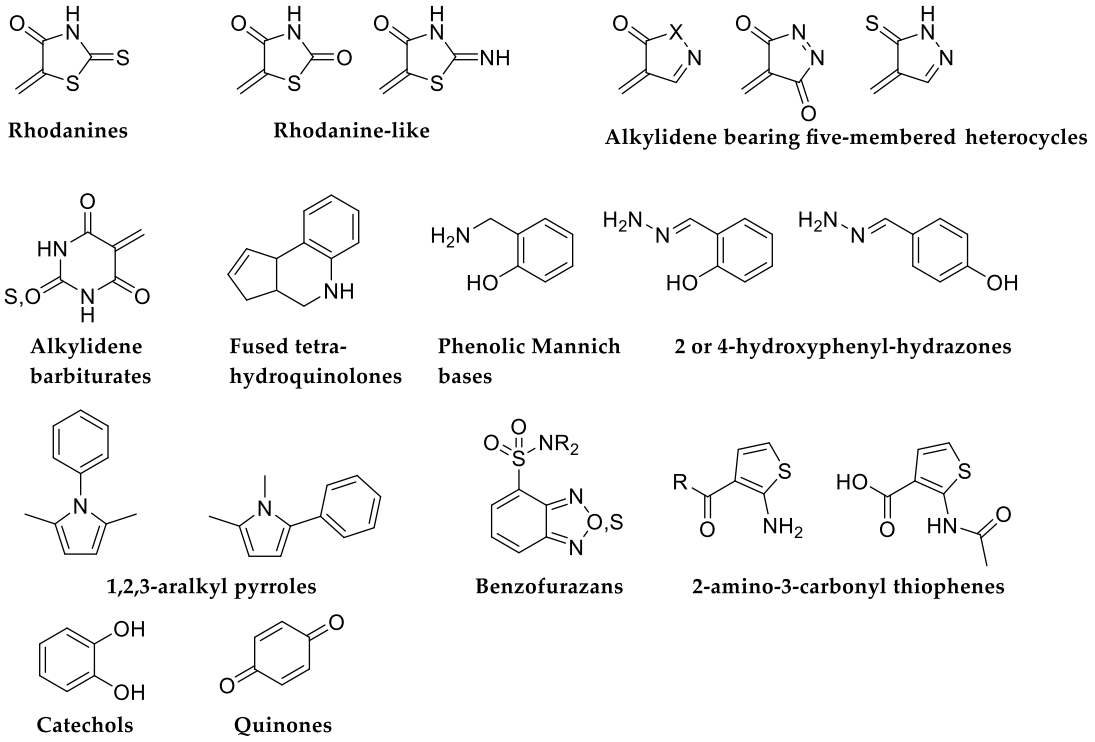


Figure 3.2: Chemical structures of pan-assay interference compounds (PAINs). These moieties are identified as frequent hitters in HTS campaigns and are generally excluded in the design of the final fragment library. Adapted from Baell and Holloway.[148]

3.1.5. Fragment Library Design Approaches

Many academic and pharmaceutical companies are intent on keeping library design methods, and indeed library content, undisclosed to maintain intellectual property. Therefore, it is difficult to compare the various strategies used between pharmaceutical groups. The only parameter that is generally published is the reported hit rate of the library. However, most libraries developed in Pharma are constructed in house by cheminformatics approaches by analysing a representative set for diversity, PAINs and solubility. The following sections summarise a few examples of library design strategies used by pharmaceutical and academic groups. The methods employed in this thesis use a combination of the approaches explored here.

3.1.5.1. Cheminformatics Filtering

The most published methods of fragment library design involve filtering of available compounds for desired physiochemical properties, the Ro3, diversity analysis and a subjective input from a medicinal chemist. For example, pharmaceutical company GlaxoSmithKline (GSK) routinely use fragment screening in their drug discovery projects. The fragment screening set they employ was developed by taking a large set of available in-house compounds and applied a set of 2D substructure and property filters to identify compounds (heavy atoms < 22, rotatable bonds < 6, donors < 3, acceptors < 8, cLogP < 2.2). A synthetic handle was also included to allow medicinal chemists to easily follow up on hits. Diversity analysis was also used to select a subset based on 3D pharmacophores to create a second fragment library.[149] Another example comes from researchers at Pfizer who describe the design of their Global Fragment Initiative (GFI), a 2,885 member fragment library that is, supposedly, broadly applicable to any target using any screening method, i.e. NMR, X-ray, SPR, MS and biochemical assays.[150] Compounds in this library were subject to rigorous filters including: reactive moieties, solubility at 1 mM in aqueous buffer and chemical complexity. Most of the fragments came from commercial sources, however 293 were synthesised specifically for the library. Interestingly, a large number of charged molecules were included (cationic and anionic) based on the observation that many approved drugs are charged. The fragments were also investigated for their 2D and 3D similarity in order to maximise diversity and ensure commercial availability of hit analogues

3.1.5.2. Target Focussed Fragment Libraries

Recently, smaller fragment libraries, directed at a specific target class have emerged as useful fragment sets for screening. One of the fastest growing target classes for drug intervention are epigenetic classes of protein, including: bromodomains, histone deacteylases and histone methyltransferases. As described in Chapter 1, these proven druggable families have led to companies, for example Enamine, creating focussed fragment libraries. In particular, Enamine, has created a PPI focussed fragment library to enrich for active molecules containing privileged core structures. This library contains compounds with similar core structures used for identification of new hits, however, as can be seen from table 3.1, does not conform to the Ro3.

Parameter	Range
Molecular mass	200 – 400 Da
cLogP	0 – 4
Hydrogen bond acceptors	0 – 5
Hydrogen bond donors	0 – 4
Rotational bonds	0 – 5

Table 3.1: Table describing the criteria used for selection of PPI fragments in the Enamine PPI fragment collection. Adapted from Enamine.[151]

Protein kinases are another example of an important drug target class for therapeutic intervention. An in-depth analysis of kinase inhibitors by Akritopulou-Zanze and Hadjuk led to the design of new kinase focussed libraries. The authors developed a scaffold-based approach that identified a novel class of low molecular mass compounds with kinase activity. This analysis of known kinase inhibitors would allow academics and pharmaceutical companies to design novel fragment collections to either increase intellectual property freedom or increase chemical space coverage.[152]

3.1.5.3. Fluorine Fragment Libraries

As described previously (Chapter 1.4.3), ^{19}F ligand-observed NMR screening is currently one of the most popular methods of screening fragment libraries. ^{19}F NMR spectra are much simpler than ^{1}H spectra as there are no background signals from endogenous protein signals. However, a caveat of ^{19}F NMR screening is that it requires fluorinated fragments in the library. This introduces a significant bias in the design of the library owing to constraints of the assay conditions. The main concern with this restriction is the limiting of chemical space to fluorinated fragments. Jordan *et al* carried out a comprehensive analysis of the coverage of fragment space of fluorine containing molecules. Using a total of 67,000 reference compounds and selection filters, diversity analysis of ^{19}F fragment space was assessed. Figure 3.3 shows a graphical representation of the coverage of fragment space by fluorine containing molecules. The data is analysed by principal component analysis (PCA) using the selection filters that were applied to define the fragment space. PCA is a method to reduce the number of dimensions in a dataset, such that the variation in the lower dimensional data is maximised. The authors conclude that the overall coverage of fragment space with fluorine containing

molecules is approximately 30%, however, the diversity is sufficient enough to be used for screening.

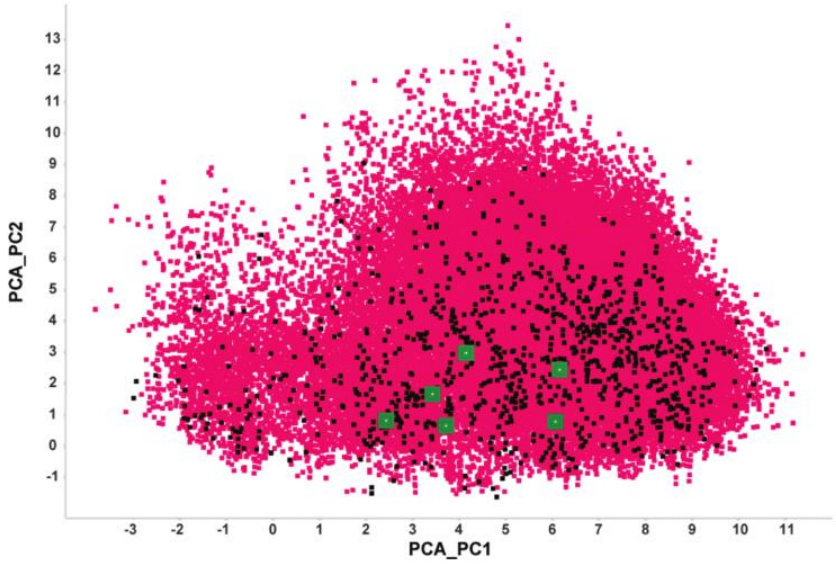


Figure 3.3: A graphical representation of fragment space coverage by ^{19}F compounds. Red dots represent compounds contained in fragment space, while black dots represent compounds in the fluorine fragment library. Green squares are hit compounds from a screen β-secretase described in the paper.[153] PCA_PC1 and PCA_PC2 are the principal components that describe each compound in the reference library.

A novel fluorine library for NMR screening is described by Dalvit *et al* who generates a library of fluorine containing fragments based on the local environment around the fluorine nucleus.[154] Fluorine atoms with different electronic charge distributions (i.e. different chemical shifts) will interact differently with the receptor, they call this the 'Local Environment of Fluorine' (LEF) library. In-house and commercial compounds were clustered according to a new fluorine fingerprint that clusters fragments based on LEF. They were also clustered using the whole molecular structure, using FCFP4 fingerprints (described in section 3.1.6.3), the two criteria give very different results as illustrated in table 3.2. This approach to library design represents a strategy directed towards the identification of fluorophilic protein environments.

Overall, the general approach to library design often depends on available resources and the intended fragment screening method. The methods described in this thesis focus on ligand-observed ^{19}F NMR as the preferred biophysical screening method. As shown previously, using fluorine containing compounds limits the chemical space coverage, however the high sensitivity of NMR is considered to supersede this limitation.

The following sections addresses the issues of fluorine fragment library design and the strategies employed in this thesis towards designing an in house fluorinated fragment library.

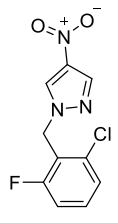
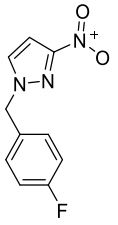
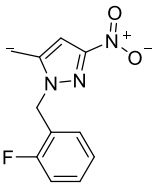
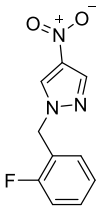
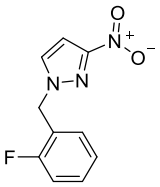
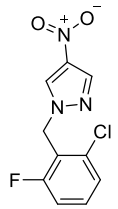
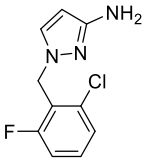
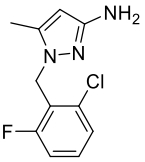
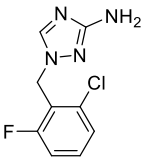
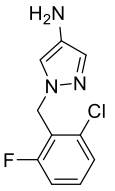
Clustered according to FCFP_4 Fingerprints					
A					
F-cluster	**230**	**2**	**31**	**31**	**31**
Global cluster	**205**	**205**	**205**	**205**	**205**
Clustered according to LEF					
B					
F-cluster	**230**	**230**	**230**	**230**	**230**
Global cluster	**205**	**126**	**126**	**77**	**58**

Table 3.2: Fluorine fragments clustered by two different descriptions. A) The five molecules belong to the same global cluster, 205, but to different F-clusters generated using the LEF fingerprint. B) The five molecules belong to the same F-cluster, 230, but to different global clusters. Adapted from Dalvit *et al*.[154]

3.1.6. Cheminformatics

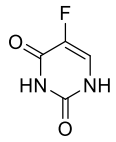
The library design strategies described above were adapted and employed here to design an in house fluorinated fragment library. This chapter relies on a range of cheminformatics methods, here I introduce the basic methods and terminology used to handle the data.

Cheminformatics was a term coined by Dr. Frank Brown in 1998 who described it as: *"… the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization."* The main focus of cheminformatics is the handling of data

related to storage, indexing and searching of chemical compounds. Here, it is a term used to describe the range of methods used in designing fragment libraries. The main aspects of cheminformatics, with regards to fragment library design, include representing chemical structures and the calculation of molecular properties.

3.1.6.1. Representation of 2D Structures

There are various ways to computationally represent and handle chemical structures and their molecular properties. For example, the cancer drug 5-Fluorouracil can be illustrated in an image file (Figure 3.4, **1**) or described by its IUPAC name (International Union of Pure and Applied Chemistry) (Figure 3.4, **4**).

1. 

2. 5-Fluorouracil
3. FC(C(N1)=O)=CNC1=O
4. 5-Fluoro-1H,3H-pyrimidine-2,4-dione
5. 1S/C4H3FN2O2/c5-2-1-6-4(9)7-3(2)8/h1H,(H2,6,7,8,9)

Figure 3.4. Chemical representation of 5-Fluorouracil (5-FU). The molecule 5-FU is represented as: 1. Chemical structure image file; 2. Systematic name; 3. SMILES; 4. IUPAC name; 5. InChI.

However, to perform computational analyses, such as molecular property prediction and substructure searches, these notations are impractical. A popular method for describing the connections of atoms in a molecule is the use of SMILES (Simplified Molecular Line Entry System) which constitutes a linear notation system (Figure 3.4, **3**). To express a molecule as a string of atomic connections, SMILES has various rules including: atoms are represented as their atomic symbol, several symbols represent bonds and their character (e.g. = stands for double bond), branches are indicated with brackets and upper and lower case letters express aliphatic and aromatic atoms respectively. InChI (International Chemical Identifier) is also another method of representation in which every structure is described by a unique string composed of layers of information. These layers encode information pertaining to atom connectivity, tautomeric information, charge and stereochemistry.[155] Figure 3.4, **5**). For the following work, SMILES strings are used as the predominant molecular descriptor for property prediction and database searching.

3.1.6.2. Pipeline Pilot Program

Pipeline Pilot is a cheminformatics workflow tool that offers an array of scripts for creating and analysing large, complex data sets, including cheminformatics, sequence analysis and modelling. It automates data analysis processes to allow fast, efficient work flows for analysing and reporting data. There are many opportunities to create custom protocols for specific functions and therefore adapt the scripts to allow focus on innovation. The following method for fragment library design utilises a clustering algorithm implemented with the program Pipeline Pilot from Accelrys.

3.1.6.3. Clustering

Clustering, or cluster analysis, is a method used to group a set of objects, based on some definition of similarity and pertaining to a defined threshold. It works by placing members of a dataset so that the distance (defined by some similarity metric) between members is minimised. Objects contained within the same group (cluster) are therefore more similar to each other, than those contained in another cluster. Clustering, itself, is widely used across an array of disciplines including cheminformatics, machine learning and image analysis. There are many different models used to define clusters, which in turn use specific algorithms that differ significantly in their definition of clusters and how to efficiently find them. The approach used by the program Pipeline Pilot uses a non-hierarchical method to cluster data sets. A set of compounds, from an input library, is divided into smaller sub-sets based on the maximum dissimilarity between compounds. The similarity score is determined using the 2D fingerprints 'FCFP4' (functional connectivity fingerprints of radius 4). This is a fingerprint of the molecule designed specifically for similarity searching.[156] For instance, a random molecule is chosen as the first cluster centre, then the most distant molecule to this one is the next cluster centre. This is repeated until there are sufficient cluster centres described. The remaining molecules from the dataset are then distributed between these cluster centres to create the final cluster.

3.1.6.4. Molecular Property Prediction - cLogS

The molecular solubility is a critical property of a drug compound as this affects the distribution across the human body. The drug needs to be water soluble enough to be transported throughout the body, yet hydrophobic enough to pass through cell membranes. A common measure of a molecule's solubility is the LogP, which is a measure of the distribution between octanol and water and gives information on the compounds

lipophilicity. However, the aqueous solubility, LogS, affects the overall bioavailability i.e. uptake and distribution, of the compound in their effective active site. For FBS we are interested in the aqueous solubility of fragments for screening as this is critical to the assay conditions. Therefore, predicting the molecular properties, such as logP/S/D, pKa and analytical spectra (NMR, mass spectrometry etc.) of compounds is a powerful tool to the computational chemist. There are currently a large number of scientific algorithms available to tackle this problem, however, these are usually developed using multiple computer languages and are difficult to transfer between different laboratories.[157] The virtual computational chemistry laboratory (VCC) is a multi-platform, online data analysis tool that allows one to perform a series of comprehensive molecular property predictions online.[158] The specific software package described here is the ALOGPS 2.1 Java applet used to predict the aqueous solubility of compound fragments needed for library design and screening. Compounds can be uploaded easily in SMILES format and the output is very fast and efficient.

Fragment based screening is in widespread use, hence fragment library design is a major topic. Many review articles exist on the ways and means of fragment library design that share knowledge on the importance of the principles described here. The following aims are intended to build on these strategies for application in the following chapters.

3.1.7. Aims of this chapter

The aim of this chapter is to use three different approaches to design low-cost, fluorine-based fragment libraries that can be used for screening by ^{19}F ligand-observed NMR. A fluorine fragment library using the cheminformatics methods and design principles described will be generated and compounds purchased to include in the screening collection. The libraries will be profiled and screened for novel, selective binders against a number of validated targets considered important in the Auer group. The three approaches encompass different strategies towards library design, and it is hypothesised that using these different strategies in parallel will provide a greater chance of success.

In the first approach fragments are selected from the ChEMBL input database of known active compounds. Active compounds are filtered according to desired physiochemical properties and selected fragment sets are further reduced according to cost. It is hypothesised that known active compounds may provide a useful starting point for development into a selective ligand. Owing to the small size of fragments, selectivity is generally an issue therefore optimisation

of fragments is possible. This set is referred to as the 'Recognised Active Fragment Library' (RAFL).

The second approach uses a conventional method to cluster fragments based on their 2D similarity. The input library of compounds is the online database of commercially available compounds eMolecules. This large, online database is filtered for desired fragment properties then clustered using the Pipeline Pilot clustering algorithm FCFP4. Fragments from clusters are chosen to maximise diversity, reduce reactivity and reduce the final cost, to follow the remit of a low-cost fragment library. This set is referred to as the 'Clustered Fragment Library' (CFL).

The third approach introduces a new measure of diversity to identify a fragment set that gains the maximal diversity within that set. Due to the availability, good quality and diversity of fluorinated compounds, Fluorochem was chosen as the optimal vendor for purchasing fragments The input library of fragments is filtered according to the desired properties of the fragment library. A new measure of diversity is introduced that allows one to cherry pick a set of compounds to maximise the diversity. This follows again the objective of this project to gain a highly diverse, yet low-cost fragment set that will provide hits to our protein targets. This set is referred to as the 'Directed Diversity Fragment Library' (DDFL).

Also, a short analysis of the physiochemical properties and diversity of commercially available fragment libraries is presented.

3.2. Results and Discussion

3.2.1. A Commercial Library Analysis

There is an ever increasing number of commercial companies offering fragment collections for screening purposes. As one might anticipate, this would lead to a significant overlap of chemical space coverage between various libraries. A disadvantage of using commercial fragment libraries, and sourcing compounds from commercial vendors, is that direct competition results from sampling the same chemical space. Compound libraries were provided by commercial vendors and were supplied as structure-data file (SDF) files, allowing fast and easy similarity searches and molecular property prediction. We were able to obtain SDF files for 13 commercial vendors, listed in table 3.3, and apply common Ro3 filters to ascertain the quality of the library. Again, this is a rather subjective view, as the Ro3 is known

to be a rough guideline, however, it allows us to realise which vendor is selling the highest quality library for our requirements.

3.2.1.1. Classifying Fragments

The main aim here was to determine the quality of commercial fragment libraries, and how applicable they would be to an NMR screening approach. Firstly, we defined some simple filters to check how each library conforms to a reasonable fragment set. These were: removal of PAINs compounds, a molecular mass less than 300 Da and an aqueous solubility of greater than 100 μM (cLogS > -4). Solubility was calculated using the ALOGPS 2.1 cheminformatics lab described in section 3.1.6.4. Immediately from table 3.3 it was evident that the Maybridge (part of Thermofisher) library loses over 70% of fragments that do not conform to simple filters. Particularly alarming is the aqueous solubility filter which accounted for 41% of loss of fragments. This library would not be useful for NMR solution screening owing to the high solubility requirements for detection (> 100 μM). Figure 3.5 illustrates the distribution of compounds within the Maybridge fragment set before application of filters.

Supplier	Original Count	No PAINS	Mw <300	cLogS > -4	Final Count	Loss %
Analyticon	30	-	-2	-1	27	10
Asinex	2,284	-70	-	-396	1,818	20.4
Bionet	1,952	-50	-6	-285	1,611	17.5
ChemDIV	380	-4	-2	-32	342	10
Enamine	6,842	-20	-	-1336	5,486	19.8
Life Chemicals	2,253	-1	-	-621	1,631	27.6
Life Chemicals Pooled	1,010	-	-	-238	772	23.6
Maybridge	5,244	-194	-2420	-1090	1,540	70.6
Otava	1,217	-22	-8	-169	1,018	16.4
Prestwick	99	-3	-	-	64	35.4
TimTec	91	-3	-	-28	60	34.1
Vitas-M	2,298	-113	-	-721	1,464	36.3

Table 3.3: Commercial library analysis. Table showing the percentage loss of each commercial library when filters are applied. The extraction of information and cheminformatics processing illustrated in this table was performed by Dr Steven Shave.

On average, there is a 20% loss of fragments from each library that do not conform to 'basic' filter sets.

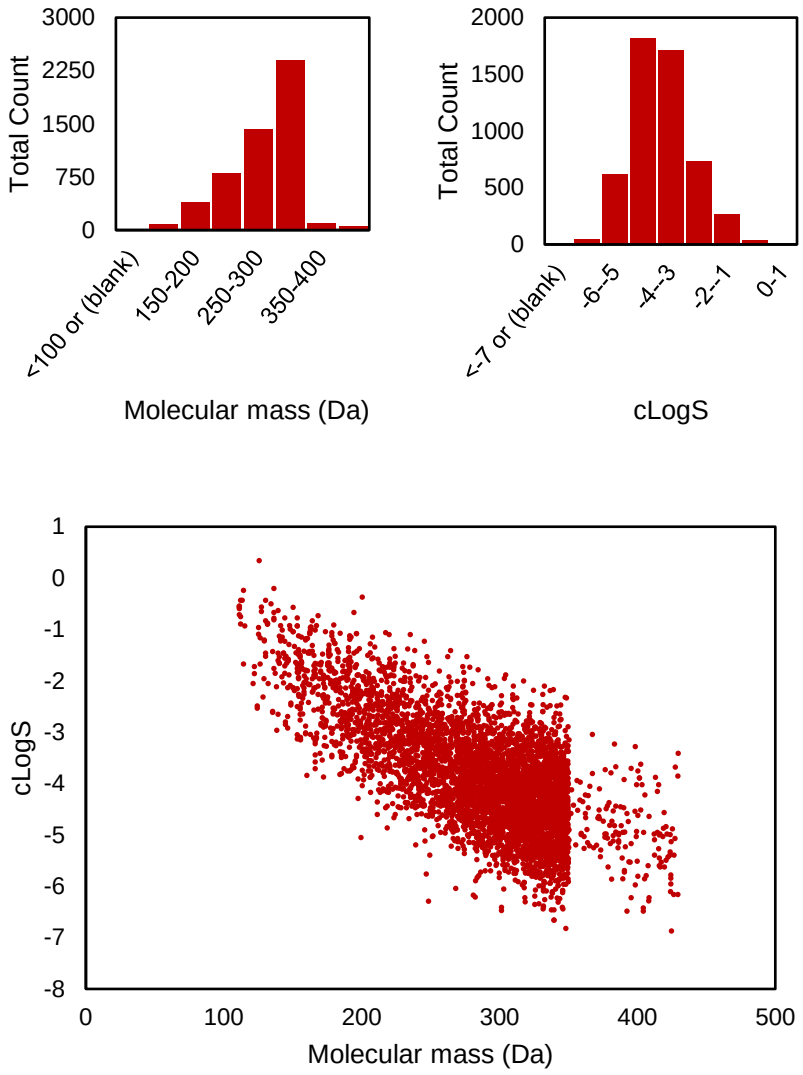


Figure 3.5. Maybridge fragment screening collection statistics. Top Left: Plot showing the frequency distribution of the molecular mass of fragments present in the library. Top Right: the frequency distribution of the aqueous solubility of fragments present in the library. Bottom: A 2D plot of aqueous solubility (cLogS) vs molecular mass of fragments in the Maybridge collection. Applying simple, common filters to the Maybridge collection reveals that it is not Ro3 compliant and would not be very suitable for NMR-based screening

Before deciding on which library to purchase, another aspect to consider was the coverage of fragment space. One might expect a significant overlap of coverage space between commercial libraries owing to the small volume of fragment space and the large numbers of compounds available in libraries. However, the fragment crossover between libraries was surprisingly low as illustrated in table 3.4. This was done by using each library as a 'query' and cross-referencing compounds to every other library. For example, there are 10.86% of fragments present in the Maybridge library that are also present in the Bionet library. The average

crossover was 2.6 % which implies that there are a lot of unique fragments present in each set. This is an advantage for consumers as this reduces competition if hits from screening are identified.

An important objective from designing our in-house libraries was to reduce the cost as much as possible, whilst maintaining a maximal level of diversity. The aim is to investigate different strategies to increase diversity as a function of compound number. Commercial libraries containing > 1000 compounds are on average $7 (£5) per compound, however, a usually a minimum order of > 1200 compounds (£6000), this data is taken directly from company websites. Usually, ordering more compounds reduces the price per compound. Also, a disadvantage of commercial libraries is a general trend of vendors is to supply compounds as 1 – 10 mM DMSO stocks or 1 mg powder in micro-well plate form. This is unsuitable for an NMR screen where higher stock concentrations are needed due to the large volumes required for NMR.

Query	Analyticon (30)	Asinex (2284)	Bionet (1952)	Chem DIV (380)	Enamine (6842)	Life Chemicals (2253)	Life ChemicalsPooled (1010)	Maybridge (5244)	Otava (1217)	Prestwick (99)	TimTec (91)	Vitas-M (2298)
Analyticon (30)	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Asinex (2284)	0.00	100.00	2.58	1.01	5.34	3.77	2.41	1.84	4.73	0.22	0.48	13.84
Bionet (1952)	0.00	2.97	100.00	0.10	5.89	1.95	1.54	10.86	4.25	0.87	0.20	4.71
Chem DIV (380)	0.00	6.05	0.53	100.00	0.26	1.84	1.32	0.26	1.58	0.26	0.26	3.16
Enamine (6842)	0.00	1.74	1.72	0.01	100.00	1.02	0.64	1.32	1.89	0.09	0.03	2.44
Life Chemicals (2253)	0.00	3.91	1.60	0.31	3.11	100.00	36.66	1.20	4.35	0.04	0.04	2.93
Life Chemicals Pooled (1010)	0.00	5.45	2.97	0.50	4.36	81.78	100.00	1.19	4.65	0.00	0.10	2.67
Maybridge (5244)	0.00	0.80	4.04	0.02	1.72	0.51	0.23	100.00	1.37	0.38	0.08	1.95
Otava (1217)	0.00	9.04	6.57	0.49	10.93	8.22	3.86	5.92	100.00	0.49	0.16	12.00
Prestwick (99)	0.00	5.05	17.17	1.01	6.06	1.01	0.00	20.20	6.06	100.00	1.01	6.06
TimTec (91)	0.00	10.99	4.40	1.10	2.20	1.10	1.10	4.40	2.20	1.10	100.00	7.69
Vitas-M (2298)	0.00	13.75	3.87	0.52	7.22	2.96	1.17	4.44	6.35	0.26	0.30	100.00

Table 3.4: Fragment library crossover table. This table illustrates the percentage of the query library (left column) that is present within the criteria adherent candidate library (top row). The crossover is given as a percentage of the query library present in the candidate library. The analysis and extraction was performed by Dr Steven Shave

3.2.2. Defining Physiochemical Property Filters

To design our own, in-house fluorinated fragment libraries it was necessary to define the physiochemical properties that are required for NMR screening. Figure 3.6 illustrates the desirable properties for fragments in our library to possess.

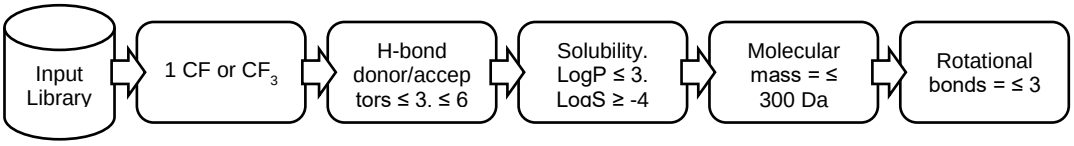


Figure 3.6: Pipeline of desirable physiochemical properties to be used as filters to define in-house CTB fragment sets. The input library of compounds was filtered according to each box in the workflow. The input libraries are described in the main text for the different design approaches.

The Ro3 states a molecular mass less than 300 Da should be used as a guideline for fragments. As illustrated in table 3.2.1, most of the libraries conform to this rule with the exception of the Maybridge library. This was chosen as the molecular mass cut off for CTB fragments. Aqueous solubility is a critical parameter for NMR fragment-based screening. A solubility threshold is a good selection filter as this tends to remove many compounds and is dependent on the type of assay to be applied for screening. An aqueous solubility of 100 μM is required for detection of ^{19}F in NMR (see chapter 4.2.2), this equates to a cLogS of > - 4, therefore this parameter was also selected as a filter. As described previously (Section 3.1.4) primary screening hits must be scrutinised for possible artefacts, i.e. it is necessary to confirm that their activity is dependent on a specific interaction. Pan-assay interference compounds (PAINs) are defined as compounds that contain specific chemical structures that show activity across a range of assays and against a range of proteins. Therefore, including a PAINs filter in development of the libraries was necessary to remove any potentially promiscuous compounds. Hydrogen bond donor and acceptor pairs and number of rotational bonds were used as additional filters to reduce the size of fragments returned during the search of chemical space.

The most important filter used in the pipeline workflow is the fluorine atom filter. The fluorine atom needs to be included in each fragment for ligand binding detection in the NMR screening assay. The benefit of fluorine NMR screening is that compounds with a single CF or CF_3 group will have a single peak in the NMR spectrum, i.e. one singlet resonance per compound. This makes deconvolution and interpretation much easier as opposed to including compounds

with multiple magnetically distinct fluorine atoms, therefore, for the purposes of screening, it is desirable to keep fluorine nuclei to a minimum Also, it has been shown that CF_3 or OCF_3 moieties always reduce the aqueous solubility, whereas single CF groups can alter it in either direction.[159] . The single F or CF_3 filter was therefore applied to make spectral analysis straightforward.

Other requirements in the selection of compounds include removing salt forms of fragments that are prone to precipitation after multiple freeze thaw cycles. Also, a rather subjective medicinal chemistry viewpoint on desirable chemical functionalities and 'interesting' looking compounds was used to select fragments. The following sections describe the results of library generation by each of the three methods of library design.

3.2.3. Recognised Active Fragment Library (RAFL) (Performed with Dr Steven Shave).

The first approach to constructing a fragment library was built on the hypothesis that known active fragments in the literature, that hit more than one target, may be a useful starting point for lead generation. These recognised actives may have alleged promiscuous activity, however, at such low molecular weight, optimisation into selective compounds is not uncommon. ChEMBL is an online database containing known, bioactive, small molecules. As of May 2017, there were 1,735,442 active compounds (of which 1,727,112 have .mol files). The data is curated from the available scientific literature and it covers parameters including: 2D structure, LogP, molecular weight, binding constants, pharmacology and ADMET data.

Using the parameters described in section 3.2.2 to define the CTB fragment set, the fragments according to these filters were extracted from the ChEMBL database. This resulted in 30 known fragment actives that had: M_w < 300 Da, cLogS > 4, 1F or CF_3 containing moiety (example fragment in figure 3.7). This is only 3% of the available 'fragment space' included in ChEMBL; that is 960 fragments were extracted without the 'fluorine atom' filter (Figure 3.8). Also, 29 of 30 compounds only occurred once in ChEMBL, i.e. they were only active against one target. This could infer some selectivity, however, many of these compounds were found as part of the SAR exploration series and not part of a fragment screen. Therefore, their activity against one target is perhaps not surprising. For example, compound RAFL08 (figure 3.7) was investigated as a new substrate for semicarbazide-sensitive amine oxidase enzyme.[160]

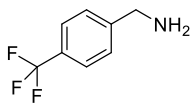


Figure 3.7: Chemical structure of fragment RAFL08.

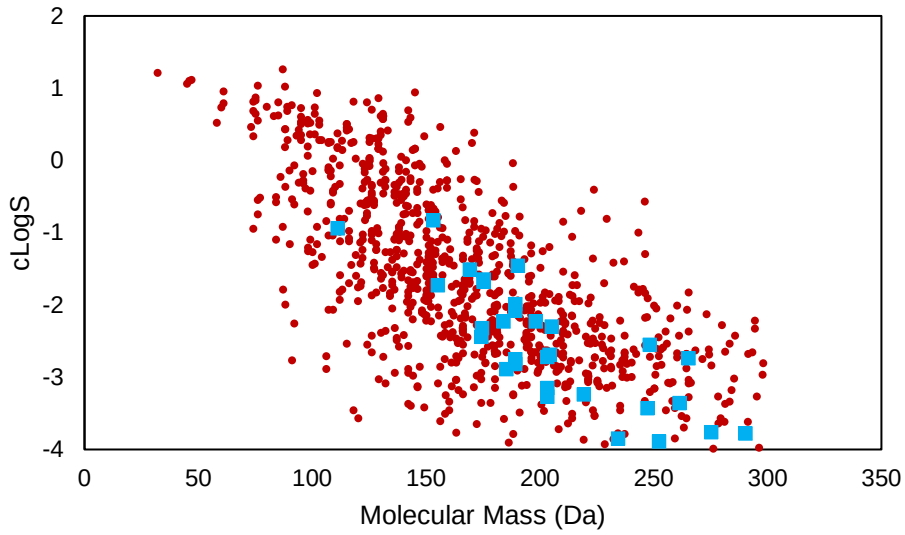


Figure 3.8: A plot illustrating the ChEMBL fragment space in two dimensions: molecular mass and aqueous solubility. The red circles represent non-Fluorine containing compounds extracted from ChEMBL The blue squares represent the fluorine containing compounds.

Also, pertaining to our objective of producing a quality, diverse library as a function of total cost, each compound was priced from the cheapest vendor, for 30 compounds the price was £3197. This was deemed out of budget, therefore a price per compound cut-off filter of £120 was applied. This reduced the number of compounds to 22 and a final library value of £992, an average of £45.10 per compound. As many companies require a minimum weight order for custom syntheses, the benefit is that this provided enough material of each compound (~ 1 g) for chemical follow up of any hits. The disadvantage of 'cherry picking' compounds is that it is more expensive. These compounds were purchased and their structure and quality check data is presented in appendix 8.2 and 8.3 respectively.

3.2.4. Clustered Fragment Library (CFL)

The second method of library generation used a popular clustering method, as described in section 3.1.6, using the Pipeline Pilot algorithm. The input library of fragment compounds was eMolecules (September 2016), which is an online chemical database containing over 8 million compounds that are purchasable from commercial suppliers. Applying the filter sets illustrated in figure 3.6, reduced the number of compounds to 2625 unique fragments. Pipeline Pilot uses FCFP4 (Functional Connectivity Fingerprints of radius 4, see section 3.1.6.3), to represent a chemical structure to assist in sub-structure and similarity searching.[161] We pre-

defined the number of clusters to 200 cluster centres, and fragments were placed into each cluster based on their FCFP4 score. In selecting compounds from clusters, the goal is to cover as much chemical space as possible, i.e. ideally one compound from each cluster. The output is a list of clusters containing compounds with the centroid as the first compound in the list, therefore priority was given to this compound (example: circled compound in figure 3.9). This ensures that the final selection has the highest diversity. However, this was not always possible with the goal of producing a low cost library as there were several compounds more expensive compared to the threshold of price per compound. Therefore, the compounds priced within the CTB threshold, from as many clusters as possible were priced to gain maximum diversity and minimum cost. Also, as compounds were taken from a commercial database, there were also salt forms of compounds present. These are undesirable as they may be unstable after storage in solution. This resulted in 24 compounds being selected, however, 4 of these were from the same cluster, skewing the diversity. The overall price for 24 compounds was £607, which was deemed reasonable for the amount provided.

It is often easier to purchase compounds than synthesise them, therefore choosing compounds from clusters containing more than one compound might represent an effective approach. This is because fragment analogues can be easily purchased from the same cluster to allow rapid screening and SAR analysis.

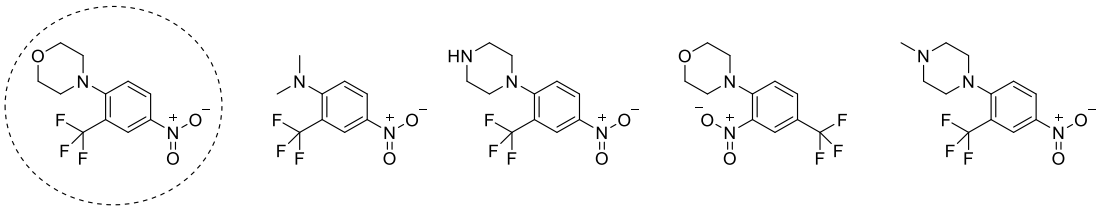


Figure 3.9: An example cluster output from the clustering protocol implemented in Pipeline Pilot. The circled compound illustrates the centroid of the cluster. From left to right illustrates the compounds in order of decreasing similarity within the cluster.

The final library contained 24 compounds, each with a single CF or CF_3 group, molecular mass under 300 Da, H-bond D/A > 3 and 6 respectively. Pipeline pilot uses an in built algorithm to calculate aqueous solubility, however, purchased compounds were double checked with ALOGPS 2.1 (see section 3.1.6.4) and were also found to be inside the limit. Compound structures, data and quality check analysis are illustrated in appendix 8.2 and 8.3 respectively.

3.2.5. Directed Diversity Fragment Library (DDFL)

3.2.5.1. Covering Fragment Space: A new measure of diversity

The cheminformatics work described here was completed by Dr. Steven Shave. The programming and calculation of A2A distances was also performed by Steven Shave. The input from myself was to aid in design of the algorithm in which to select the best compounds. For example, defining desirable properties, chemical functional groups and sensible chemical structures of fragments in the final set for optimisation of the final library.

In a model library, all commercially available fragments would be purchased and screened, i.e. this would cover the entire accessible chemical space. However, this is impractical, therefore it is necessary to create a set of fragments to gain maximum diversity over the largest volume of fragment space. Hence, it is also necessary to be able to measure and quantify the diversity of a given fragment set. To do this, compounds can be represented by their 3D shape, a fundamental property that is directly observed in electron density maps. Assuming that shape complementarity is a requirement for ligand binding, it follows that a molecule with a similar shape could have similar binding modes and affinities. Ultrafast Shape Recognition (USR) is a moment-based virtual method that uses the relative positions of bonded atoms to describe the 3D molecular shape.[162] Essentially, the molecular shape is encoded by a set of distributions (in this case 12 USR descriptors) that is unique to a set of spatial coordinates. This means that for each compound, these moments only need to be calculated once and can be stored easily in a database. USRCAT (Ultrafast Shape Recognition with CREDO Atom Types) is an extension that includes the pharmacophoric features with atom types in encoding compounds using a total of 60 USR descriptors. A pharmacophore is a description of the topological features of a molecule that contribute to the recognition of that molecule by a biological receptor. Including this information greatly enhances the power of the molecular similarity technique, pairing not only shape, but electrostatics. The similarity between compounds can therefore easily be calculated by taking the sum of squared differences between two molecules and their descriptors. This allows one to describe a simple model of diversity by a new parameter called the 'All to All Distance' (A2A). This measures the distance of every molecule in a library to every other molecule described by 60 descriptors, and hence, describes the diversity of the library (Figure 3.10).

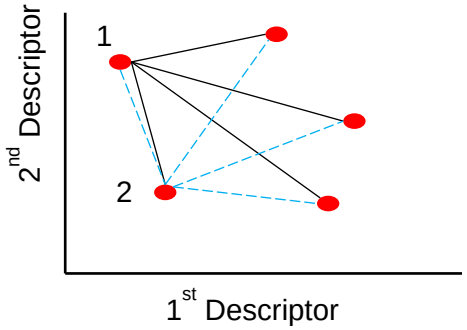


Figure 3.10: Compounds (red circles) represented as nodes with two descriptors (x and y coordinates on a 2D plane). The solid black lines indicate the distance measured from compound 1 to each other compound in the set. Dashed blue lines represent the distance measured from compound 2 to each other compound in the set. This is repeated for every compound included in the input fragment set. From this, an all-to-all distance can be calculated, measuring the diversity, or distance present between points.

This diversity analysis was then applied to the commercial libraries to calculate the diversity within these fragment collections, this is shown in table 3.5. Here, the greater the A2A score, the more diverse the library is, and therefore it is evident that we want to optimise the A2A score. To do this, we take the most remote compound from the input library population, and move it to our selected library population (Figure 3.11). Whilst the selected library population remains below the desired number of compounds (*N*), we can select from the input library to maximise the A2A score. However, this is what is known as a 'greedy algorithm' and does not always find the most optimum solution.[163]

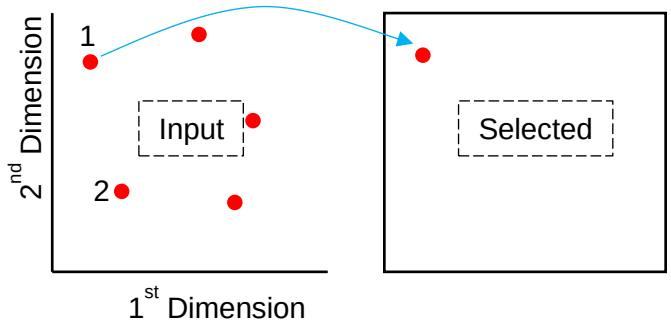


Figure 3.11: This diagram illustrates how compounds are chosen based on maximising the A2A score. The most remote of the population is taken from the input set and placed into the 'selected' library population. Whilst the 'selected' size contains less than the required number of compounds, the compound which maximises the A2A score is 'picked' from the input library and moved to the 'selected' fragment set. This is repeated until the desired number of compounds are chosen.

	Number of compounds	A2A distance	A2A/(N*(N-1))
Analyticon	27	15339.3	21.85
Asinex	1818	77110000	23.34
Bionet	1611	44938100	17.33
ChemDIV	342	2573070	22.06
Enamine	5486	5.98E+08	19.88
Life Chemicals	1631	65181800	24.52
Life Chemicals Pooled	772	14715300	24.72
Maybridge	1540	46744700	19.72
Otava	1018	21354400	20.63
Prestwick	64	75777.1	18.79
TimTec	60	62489	17.65
Vitas-M	1464	47417200	22.14
Vichem F-lib	335	3546450	31.70

Table 3.5: The diversity scores of each commercially available library by its A2A metric.

Analysing the commercial library A2A scores, if we define > 20 as a good score, is it possible to select compounds to generate a more diverse set by using all commercially available fragments. The input library was eMolecules, and a set of property filters were applied to get the best set of compounds to pick from. The filters prioritised were: M_w < 300 Da, cLogS > - 4, and single CF or CF_3 group which returned 211192 fragments (Table 3.6). The best 100 fragments were selected using the algorithm described above that maximised the A2A score. Immediately, the score appeared easy to improve upon, (Table 3.6), however this is clear as diversity must reduce with number of compounds, *N*. These compounds represent the 100 best to select from commercial fragment space. However, many of these compounds contain unusual functionalities and are expensive to purchase. We therefore filtered the commercial chemical space to a single supplier, Fluorochem. This company was chosen for its relatively large fluorinated compound set and the low cost and availability of compounds.

	Number of compounds	A2A distance	A2A/(*N**(*N*-1))
eMolecules	211192	8.94E+11	20.05
Best 100	100	6.47E+05	65.39

Table 3.6: The A2A scores of the eMolecules commercial fragment space and the best 100 compounds from this input library that maximise the A2A score.

The property filters (M_w < 300 Da, cLogS > - 4, and single CF or CF_3) were applied to the Fluorochem compounds and this returned 4934 fragments. This was reduced to 3242 after the addition of a rotational bond filter (< 3). Again, the best compounds were selected to maximise the diversity score and these are shown for different number of compounds, *N*, in Table 3.7. The output is a list of compounds in order of generating maximum diversity. Based on chemical functionalities and a price per compound cut-off of £120, 63 were selected for purchasing.

	Number of compounds	A2A distance	A2A/(N*(N-1))
Fluorochem	4934	3.81E+08	15.64
Fluorochem-best-100	100	3.11E+05	31.41
Fluorochem – filtered	3242	1.52E+08	14.43
Fluorochem – filtered	50	6.35E+04	25.92
Fluorochem – filtered	100	2.32E+05	23.42
Fluorochem – filtered	200	8.41E+05	21.12

Table 3.7: A2A scores of the fluorochem fragment set. An additional filter was included to reduce the total number of compounds; this was a rotational bond filter (NROT < 3) and removes 1692 compounds. The A2A scores from this reduced filtered set for the best 50, 100 and 200 compounds is shown in the last column.

Using this new measure of diversity, we were able to obtain the most diverse set of fragments at the lowest cost. Although fragment space coverage is very low, the diversity within our fragment set is high, giving the best chance of finding a hit compound. The details of all compound structures and quality data are given in appendix 8.2.

3.2.6. Evaluating Fragment Libraries

In summary, three libraries were designed to obtain the most diverse set of compounds at a reasonable cost. An overview of the three CTB proprietary fragment libraries is presented in table 3.8. The libraries are designed for screening with ^{19}F NMR-based methods and this is

described in chapter 4. The fragment libraries can be screened against the PPI target focus of the Auer group and 2 proteins are presented in this thesis, UbE2L3 (Chapter 4) and NusE/NusB (Chapter 5). The average molecular weights and solubility conform to the Ro3 parameter sets and the DDFL library A2A score was shown to be comparable to the commercially available libraries. This exemplifies it as a simple metric for measuring diversity and for compiling a small fragment library.

Library	Number of compounds	Average molecular weight (Da)	Average aqueous solubility (LogS)	Cost (£)	A2A
RAFL	22	189.2	-2.57	992	15.57
CFL	24	181.3	-1.79	607	16.21
DDFL	63	221.5	-2.93	4967	19.74
Total	**109**	**197.3**	**-2.43**	**6566**	

Table 3.8: Overview of the designed, in-house CTB fluorinated fragment libraries. The number of compounds in each library (before QC) and the average M_w and cLogS are given to show Ro3 compliance. The cost of each library is also shown.

Full details of compound structures and quality data are given in appendix 8.2. Quality check of compounds and pooling of the compound libraries is discussed in chapter 4.

3.3. Concluding Remarks

In conclusion, three separate methods of fragment library design were implemented in parallel to maximise the quality and diversity of selected fragment sets. The first approach was to select fragments based on the knowledge that they were active in previous biochemical assays. Fragments extracted from the ChEMBL database conforming to a set of criteria were purchased and used to construct the RAFL set of compounds.

The second approach, used a common clustering algorithm to explore the entire commercial fragment space. Fragment criteria were defined and these filters were applied to the eMolecules database of commercially available compounds. Fragments that conformed to the selected criteria were clustered according to their 2D molecular similarity. A subjective set of compounds were purchased from cluster sets to maximally represent the fragment space. One of the first steps after hit discovery is to test close analogues or larger compounds that contain the core of the fragment hit. It is always easier to purchase compounds than to synthesis them, therefore using the commercial chemical space as an input library is a useful approach. This

is sometimes referred to as a 'SAR by catalogue' approach and can rapidly increase the success of ligand discovery through easy access to commercially available, similar compounds.[164]

The third approach was introduced to increase the chances of finding a hit compound during screening. A short analysis of the available commercial fragment libraries allowed one to focus on producing a high quality, diverse set of fragments at minimum cost. The libraries were designed for primary screening by ^{19}F NMR methods. This limited the total chemical space coverage, however, a metric to ensure maximal diversity when selecting fragments was formulated. This was called the A2A score and was used to measure the diversity within commercial and selected fragment sets. This was the core principal of the DDFL library and led to the purchase of fragments that maintained high quality and diversity as a function of cost. This simple metric can also be used by academic groups looking to put together affordable, high quality and diverse libraries.

The CTB libraries are also more amenable to a wide variety of assays as the high average, aqueous solubility allows for testing at much higher concentrations of ligand. Also, the relatively low, average molecular weight of each set means that hit rates are likely to be higher, conforming to the arguments proposed by Hann *et al*.[135] The inclusion of the F atom, also means that ^{19}F NMR can be used as a secondary binding confirmation method, if primary screening is achieved by another technique.

The relatively small library sizes described here were created using established methods and shown to be far less expensive than a commercially available library. A well-established library is less challenging to work with and uses far less material throughout the course of screening. Cherry picking, i.e. selecting the best or most desirable compounds, allowed us to gain enough material for multiple assays and for chemical follow up. Although a subjective chemical view towards 'attractive' fragments introduces a selection bias, this does ensure that hit compounds have the capacity for chemical evolution.

The designed CTB libraries will be evaluated in the following chapters and were used for screening by ^{19}F NMR against a number of important PPI targets.

Chapter 4

NMR Assay Development and Evaluation of Fragment Libraries

This chapter contains the development of the NMR fragment screening experimental methods and parameters. An introduction to ^{19}F ligand-observed NMR screening techniques and relevant NMR binding kinetics are discussed. NMR screening conditions are achieved by utilising a known fluorinated ligand to human serum albumin and compound handling and pooling conditions are described. The results of the ^{19}F NMR fragment screen against UbE2L3 complete this chapter.

4.1. Introduction

4.1.1. Fluorine NMR Spectroscopy for Compound Screening

NMR spectroscopy has been the predominant method used for screening fragments since its inception by Fesik and co-workers. There are a multitude of NMR methods that are capable of detecting binding events of small molecular ligands over a very broad affinity range. Rapid technological advances in cryoprobe technologies and high-field spectrometers has led to extensive applications of NMR spectroscopy to study protein-ligand interactions within a pharmaceutical perspective.

Ligand-observed screening is ideal for NMR, advantages of this method include: small amount of protein required (nM – 10 μM), small amount of ligand required (excess ligand up to 250 μM, therefore solubility limits are good), no size limitation of protein and no isotopic labelling of protein required. Although, there are some drawbacks, namely ligands that have slow binding kinetics, i.e. high affinity (nM range), are difficult to detect. As described in chapter 1, STD, WaterLOGSY and relaxation-edited experiments are frequently used to detect ligand binding in fragment screening. The output of these experiments are 1D ^{1}H NMR spectra in which changes in intensity or sign of the respective ^{1}H signals of ligands are monitored.[165–168]

Fragment screening approaches based on ^{19}F NMR have been extensively reviewed in recent literature.[169] Fragment hits are identified by inspecting the 1D ^{19}F NMR spectrum recorded in the presence and absence of the macromolecular target. During a ligand binding event, the fluorine nucleus experiences a change in the electric fields, short range contacts and even hydrogen-bonding effects which can cause a change in intensity (line broadening) or change in the chemical shift (chemical shift perturbation, CSP) of the fluorine nucleus. The fluorine nucleus offers several advantages, for example:

- 100 % natural abundance
- Good sensitivity (0.83 % of the ^{1}H)
- Large chemical shift range
- No endogenous background signals (i.e. from buffers or proteins)

The final point above is particularly interesting as this offers possibilities to screen fragments *in vivo*. Another interesting view of ^{19}F incorporation into drug molecules is their propensity

to form hydrogen bonds. A review by Bissantz shows that interactions between CF and polar hydrogen atoms frequently occur in the PDB.[170] Another study also observed a thrombin inhibitor change its binding mode upon fluorination such that a CF---HN interaction is formed.[171]

A good example of the broad range of applicability of ^{19}F NMR methods is that of Dalvit *et al* who describe a screening method termed 3-FABS (3 Fluorine Atoms for Biochemical Screening). This method is based on the change in chemical shift of a trifluoromethylated (CF_3 containing) substrate when it is converted into a product by an enzyme. If a compound is present that inhibits the enzyme, the substrate will be converted less rapidly and the fluorine chemical shift changes will also decrease. Clearly the leader in the application of ^{19}F NMR for screening, Claudio Dalvit and co-workers also pioneered another technique, called FAXS (Fluorine chemical shift Anisotropy and Exchange for Screening) which relies on the use of a weak-medium affinity fluorinated spy molecule. The relaxation of the fluorinated molecule is monitored in the absence and presence of protein and non-fluorinated screening compounds. It is based on a competition assay, where a hit is detected if the fluorinated spy molecule has a reduction in binding. Screening compounds are then deconvoluted by individually measuring compounds from the hit pool.

^{19}F NMR can also be used for protein-observed screening approaches. Protein-observed approaches require the incorporation of a fluorinated amino acid into the primary sequence. Gee *et al* describe the small molecule screening of the CREB binding protein (CBP) (CBP/p300, KIX) PPI through protein-observed ^{19}F NMR spectroscopy (PrOF NMR). They replace native tyrosine residues with 3-fluorotyrosine and monitor the chemical shift of the fluorine resonance in the presence of screening fragments. This yielded seven final fragment hits deemed suitable after deconvolution. In this instance, a disadvantage of PrOF NMR is that deconvolution of hit fragments requires repetition and analysis of individual fragments.

In this chapter the applied experimental aspects of NMR-based fragment screening and considerations for optimum assay set up are described. In the following section the considerations for determining the ligand – protein dissociation constant, K_D, by NMR methods are explained.

4.1.2. Ligand Binding at Thermodynamic Equilibrium and Relevance to NMR Spectroscopy

NMR is a powerful technique to study the binding interactions of proteins and ligands. The characterisation of the binding properties by several NMR spectroscopic methods has been extensively reviewed.[172] Here, I will discuss the binding of the theoretically correct mathematical treatment of complexation reactions relevant to the NMR spectroscopy techniques applied in this thesis.

A simple advantage of NMR for fragment screening is that it can detect binding using compound concentrations well below the K_D of the interaction. Using low concentrations of compound is advantageous as this lowers the risk of nonspecific binding (false positives) by reducing aggregation.[173] When a small molecule ([L]) binds to a larger macromolecule ([P]) to form a complex ([PL]), the equilibrium equation, according to the law of mass action, can be written as:

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$$

Equation 4.1

The dissociation constant, K_D, is then defined as:

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[P]_f[L]_f}{[PL]}$$

Equation 4.2

Where, $[P]_f$, $[L]_f$ and [PL] are the concentrations of free protein, free ligand and protein-ligand complex respectively.[174] The rate constants, k_{off} and k_{on} describe the rate of chemical exchange between the free and bound states of the protein-ligand complex. Assuming a diffusion controlled on rate, where k_{on} is approximately 10^7 $s^{-1}.M^{-1}$, the k_{off} can be estimated according to table 4.1.

K_D	k_{off} (s^{-1})
1 mM	10000
1 μM	10
1 nM	0.01

Table 4.1: Relationship between K_D and k_{off}. Approximate off rates (k_{off}) according to K_D assuming a diffusion controlled on rate (k_{on} of 10^7 $s^{-1}.M^{-1}$).

Estimates of k_{off} are important when interpreting NMR binding experiments and there are three possible conditions under which ligand binding occurs[175]:

1) Fast exchange: here the k_{off} is greater than the difference in chemical shift between the free and bound states, i.e. a high K_D (low affinity). For a system in fast exchange, the observed NMR signal is then a weighted average of the free and bound states and one will observe a smooth transition of the signal from the free state to the bound state.
2) Intermediate exchange: k_{off} is approximately equal to the difference in chemical shift between the two states. This results in severe line broadening effects or even complete disappearance of the signal.
3) Slow exchange: where k_{off} is less than the difference in chemical shift on the NMR timescale. This results in two separate signals being observed, one corresponding to the free state, the other to the bound state.

Chemical exchange refers to the exchange between free and bound states of the ligand-protein complex. If the populations between free and bound states are unequal, then the chemical shift (δ_{obs}) observed during the fast exchange limit is a weighted average[172]:

$$\delta_{obs} = \delta_f p_f + \delta_b p_b$$

Equation 4.3

Where δ_f and δ_b are the chemical shifts in the free and bound states respectively, and p_f and p_b are the fractions in the free and bound states respectively. Any differences between δ_{obs} and δ_f therefore indicate a hit in a NMR screen. In setting up NMR experiments it is important to estimate the fractions of each component present in solution, using equation 4.2 this can be expressed as a quadratic equation with the solution:

$$[PL] = \frac{([L]_0 + [P]_0 + K_d) - \sqrt{([L]_0 + [P]_0 + K_d)^2 - 4[P]_0[L]_0}}{2}$$

Equation 4.4

Where $[L]_0$, $[P]_0$ and [PL] are total concentrations of ligand, protein and complex respectively. From equation 4.4 the fraction of protein and ligand bound can be determined by:

$$p_{Lb} = \frac{[PL]}{[L]_0} \text{ and } p_{Pb} = \frac{[PL]}{[P]_0}$$

Equation 4.5 and 4.6

Where, p_{Lb} is the fraction of ligand bound and p_{Pb} is the fraction of protein bound, these are simulated in figure 4.1 at varying dissociation constants.

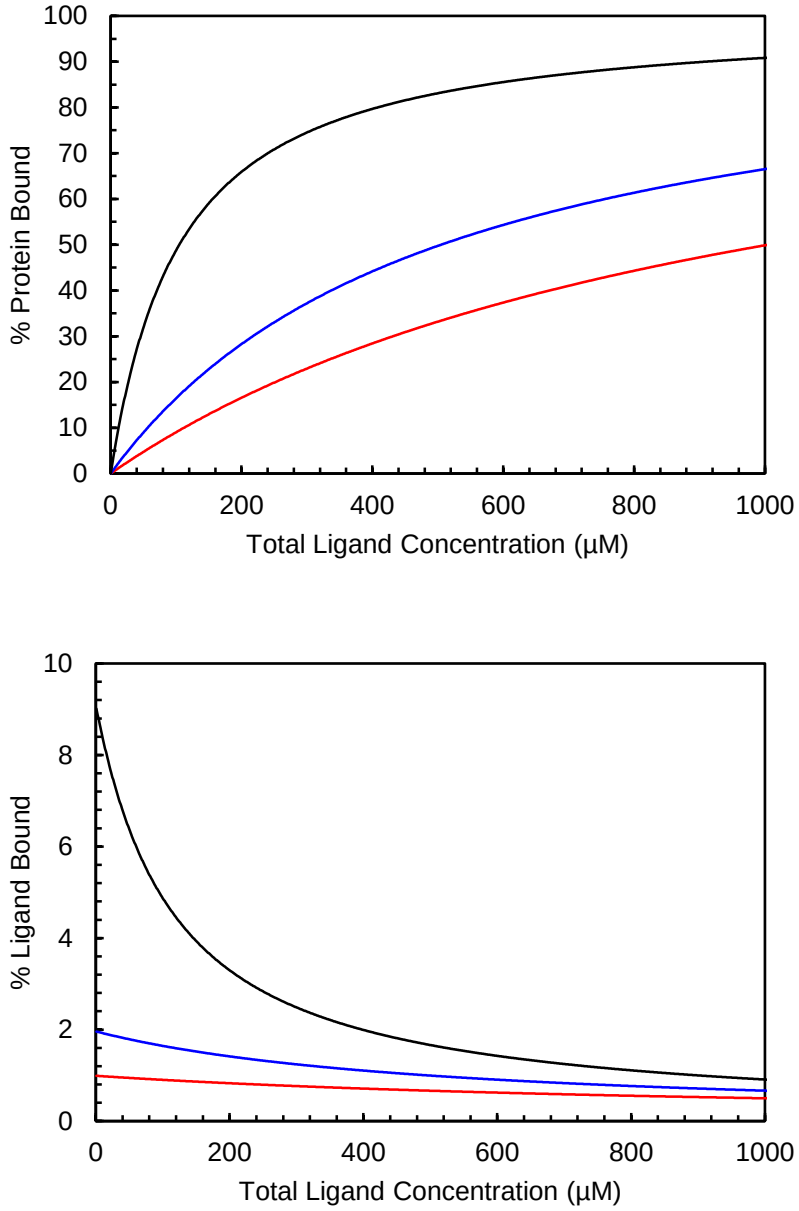


Figure 4.1: Graphical simulations of equation 4.4 to represent percentage mole fractions bound of ligand and protein at varying K_D values. $[P]_0$ is constant at 10 μM. Top: The percentage of protein bound vs ligand concentration at K_D 100 μM (black curve), 500 μM (blue curve) and 1 mM (red curve). Bottom: The percentage of ligand bound vs ligand concentration at K_D 100 μM (black curve), 500 μM (blue curve) and 1 mM (red curve). Simulations were performed using equation 4.4 and plotted in Microsoft Excel.

In the slow exchange limit, i.e. when k_{off} is slower than the chemical shift difference between free and bound states, then the free signal gradually decays, and the bound signal gradually

appears at a different chemical shift. This makes K_D determination difficult using titration experiments. Feeney *et al*. have studied the effects of intermediate exchange processes on NMR observed binding curves.[176] They describe that where the fast exchange limit is commonly assumed (equation 4.3), this can result in considerable error of up to two orders of magnitude in K_D. They acknowledge that it is essential to use ligand concentrations of the order of the protein concentration to obtain accurate K_D results. However, usually ligand-observed methods are performed in the presence of an excess of ligand. Therefore, the fraction of ligand bound is important to estimate the effect of the observed NMR parameter (e.g. chemical shift, linewidths and/or intensities) in ligand-observed NMR. The optimal concentrations to use in the screen to maximise the response is discussed in section 4.2.4. The following sections describe two methods for the determination of protein-ligand dissociation constants by NMR spectroscopy.

4.1.2.1. Differential Chemical Shift Perturbation (dCSP)

To quantify the protein-ligand interaction, the objective is to deconvolute the mole fractions of the bound and free species. Using ^{19}F ligand-observed NMR as a reporter nucleus, Jordan *et al* introduce a method called differential chemical shift perturbation (dCSP) for the determination of dissociation constants. They describe the difference between the chemical shift as:

$$\upsilon_{obs} - \upsilon_{free} = p_{Lb} \times \Delta\upsilon^{app}$$

Equation 4.7

Where ν_{obs} and ν_{free} are the frequency (in Hz) of the bound and free states, p_{Lb} is the fraction of ligand bound and $\Delta\nu^{app}$ is a complex function of the nuclear and kinetic parameters. As the ligand is in vast excess over the protein concentration ($[L]_0 >> [P]_0$) then the fraction of ligand bound is: $p_{Lb} << 1$ and $[L]_0 = [L]_f$. The fraction of ligand bound, p_{Lb}, can then be expressed as:

$$p_{Lb} = \frac{[P]_0}{K_D + [L]_0}$$

Equation 4.8

Then they describe the differential frequency shift as:

$$\gamma = \frac{\nu_{1,obs} - \nu_{free}}{\nu_{2,obs} - \nu_{free}}$$

Equation 4.9

This allows one to calculate the K_D at two different ligand concentration ($[L]_1$ and $[L]_2$) by:

$$K_D = \frac{\gamma[L]_1 - [L]_2}{1 - \gamma}$$

Equation 4.10

This is an interesting, yet novel, approach to K_D determination and the authors validated the method by comparison to SPR data in parallel[45] (see appendix 8.3.2 for derivation).

4.1.2.2. Equilibrium binding in HSQC-based NMR spectroscopy

The more frequent methods of K_D determination by NMR utilise changes in protein resonance on addition of ligand. In these experiments, ligand binding is monitored by perturbations in the ^{15}N HSQC spectra of the target protein, this is implemented in Chapter 5.

An important advantage of protein-observed NMR titrations, is that it is possible to observe the fully bound state, i.e. conditions under a fully saturated protein. This means that the maximum NMR observable parameter (e.g. chemical shift) can be measured directly. To analyse [^1H, ^{15}N]-HSQC spectra upon ligand binding requires one to weight the contributions of each nucleus to the binding event, this is defined by the average weighted chemical shift difference, $\Delta\delta_{norm}$:

$$\Delta\delta_{norm} = \sqrt{(\Delta\delta^1 H)^2 + (0.1[\Delta\delta^{15} N])^2}$$

Equation 4.11

This equation describes the contribution of each nucleus to the average distance moved by the resonance signal. The scaling factor, 0.1, in equation 4.11 has no theoretically justified value, and values have ranged from 0.1 – 0.45.[177] A threshold value of $\Delta\delta_{norm}$ = 0.04 ppm is considered to indicate that the corresponding amino acid residue is involved in the binding interaction.[178] According to Fesik *et al* the lower limit for detecting $^1H/^{15}N$ chemical shift changes is 20% occupation of the proteins binding sites.[179] Analysing figure 4.1, it can be seen that for a K_D of ~1 mM (red curve), a ligand concentration of ~250 μM ($[P]_0$ = 10 μM, 1:25 molar ratio) is

required to saturate 20% of the protein binding sites. In HSQC-based NMR screening, the total protein concentration, $[P]_0$, is lowered to reduce detection of weak affinity (> 1 mM) binders.

The dissociation constant is obtained by monitoring the chemical shift change of the backbone amide residues as a function of ligand concentration. Multiple binding curves can be determined for separate residues and the K_D value averaged over all residues that shift in the spectrum. They are then fitted by non-linear regression analysis to:

$$\Delta\delta_{norm} = \Delta\delta_{max} \frac{([L]_0+[P]_0 + K_d) - \sqrt{([L]_0+[P]_0 + K_d)^2 - 4[P]_0[L]_0}}{2[P]_0}$$

Equation 4.12

Where $\Delta\delta_{norm}$ is the observed change in the chemical shift from the free state and $\Delta\delta_{max}$ is the maximum chemical shift achieved at saturation (obtained during fitting). An extensive analysis has shown that the optimal protein concentration, $[P]_0$ is 0.5 K_D. The ligand should ideally reach a molar ratio of 10:1 (ligand:protein) to gain an accurate K_D measurement.[172,180,177]

The methods described in sections 4.1.2.1 and 4.1.2.2 are important for the analysis and interpretation of NMR data presented in chapter 5.

4.1.3. Target Selection

The protein target for screening is the ubiquitin conjugating enzyme, UbE2L3. Applying a parallel fragment approach (to the peptide building method, chapter 2) it is hypothesised to give a greater chance to discover novel ligands. The motive for targeting this protein has been described in Chapter 2.1.8.

4.1.3.1. Small Molecule E2 Conjugating Enzyme Inhibitors

The success of clinically developed proteasome inhibitors has driven awareness toward the discovery of inhibitors for the E1-E2-E3 pathways. Although small molecule inhibitors are scarce for the E2 class of enzymes, advances in screening techniques has led to the discovery of increasingly unique inhibitors of these enzymes. Table 4.2 shows a selection of E2 conjugating enzyme inhibitors demonstrating the chemical druggability of this class of enzymes.

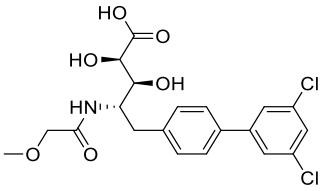
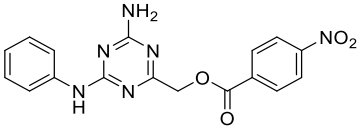
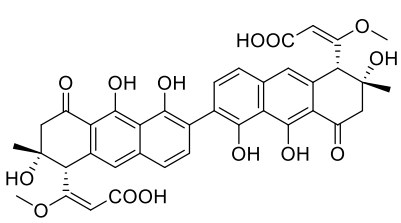
Target E2	Name	Structure	Description
UbE2R1	CC0651		Discovered from HTS for *in vitro* ubiquitination of p27. Allosteric binding to UbE2R1 causes long range conformational changes and inhibits cancer cell line proliferation.[181]
UbE2B	Triazine analogues		Pharmacophore-based virtual screening of the ZINC database led to discovery of a triazine core moiety for chemical follow up and characterisation.[182]
UbE2I	Spectomycin B1		Selectively blocks the formation of E2-SUMO thioester by binding to UbE2I. Discovered using an *in situ* sumoylation assay.[183]

Table 4.2: Examples of E2 enzyme inhibitors currently being developed for clinical use.

4.1.4. Aims of this chapter

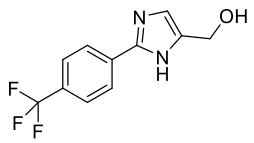
The aim of the work reported in this chapter was to develop and optimise the NMR assay parameters to the ^{19}F-FBS screening process and evaluate our fragment libraries against the ubiquitin conjugating enzyme UbE2L3. Whilst it is relatively simple to start screening fragments, ultimately there are practical considerations that must be taken into account to deliver success. These include: choosing the correct NMR experiment, fragment library design (see Chapter 3) and target selection. Also, experimental considerations of ligand-observed, fluorine fragment-based NMR screening include: how many compounds per pool, CF vs CF_3 group, protein:ligand ratios, DMSO concentration and quality control of fragments. The relevant conditions and optimisation of these parameters are described herein.

4.2. Results and Discussion

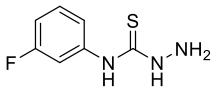
4.2.1. Preparing Fragment Stocks

The first consideration for beginning a fragment screen is to prepare suitable fragments stocks so that compounds can be easily stored and distributed into respective pools. Stock preparation is also a rudimentary measure of poorly behaving compounds, such as aggregators or insoluble compounds. The fragments were purchased from commercial vendors (see chapter 7.3.1) in sufficient amounts (mostly > 50 mg) to test solubility and to allow running of sufficient screens. The purchased amount would also provide enough material for any biochemical follow up of hit compounds. Also, compounds that were available as salts were avoided as these have a higher propensity to precipitate from DMSO after multiple freeze-thaw cycles.[184]

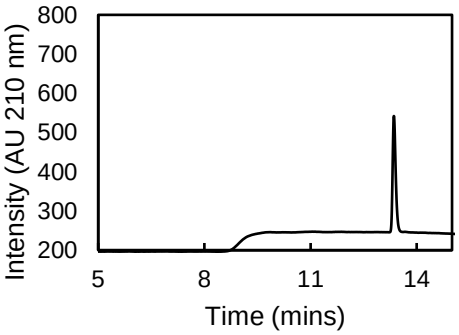
Fragments were stored in 100 % DMSO at either 10 or 100 mM concentration depending on solubility (see chapter 7.3.1) and stored at -20 °C. Fragments are stored as individual compounds, and not as pools for screening, as this reduces the probability of fragments reacting over time. Fragment concentration was determined by weight and the calculated amount of non-deuterated DMSO was added. Solubility was measured by visual inspection and all compounds were soluble at these concentrations, even after multiple freeze-thaw cycles. The stock concentration was determined by the desired concentration in the fragment screen (100 μM, see section 4.22). For example, screening a mixture of 10 compounds at 100 μM (500 μl volume), with a maximum of 1% DMSO concentration, requires individual compound stock concentrations of 100 mM. Fragment samples were also diluted to 100 μM in 25% acetonitrile:water (ACN:H_2O) for quality check (QC) of purity and identity by HPLC and LCMS (0.1% final DMSO concentration). Fragments were labelled according to the library they originated from and the ID within that library, for example, DDFL30: Directed Diversity Fragment Library, compound number 30. Figure 4.2 shows an example of QC analysis for compound DDFL30 which passed QC criteria and compound RAFL03 which failed and was not included in the final screen.



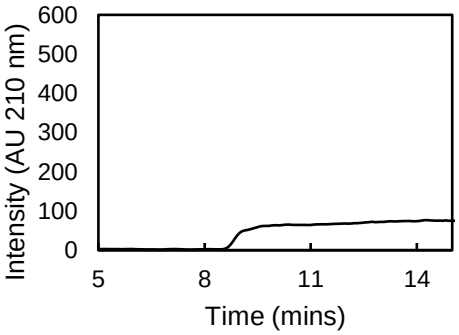
DDFL30



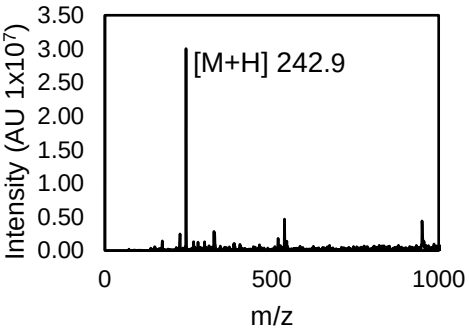
RAFL03



HPLC: t_R = 13.3 mins



HPLC: t_R = N/A



MS (ESI): $C_{11}H_9F_3N_2O$. $[M+H]^+_{calc}$ 243.1; $[M+H]^+_{obs}$ 242.9

N/A

MS (ESI): no peak observable by LCMS

Figure 4.2: An example of QC analysis for compound DDFL30 and RAFL03. Concentration of compound is 100 μM in 25% ACN/H_2O (0.1% DMSO). Top: HPLC chromatogram, R_t = 13.3 minutes, > 95% purity determined by peak area. Bottom: LC/MS spectrum, $[M+H]^+_{calc}$ = 243.1. $[M+H]^+_{obs}$ = 242.9.

4.2.2. Reference NMR spectra and compound pooling

As the NMR method used for screening is a fluorine ligand-observed method, ^{19}F reference spectra of individual compounds were recorded. Ideally, the NMR screen is performed under the same buffer conditions that the target protein is stable in, therefore, reference spectra were collected in a common biological buffer: 20 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 100 mM NaCl, 10% D_2O at pH 7.5. One dimensional ^{19}F reference spectra were acquired for each sample at 300 K on a Bruker Avance 400 MHz

spectrometer equipped with a BBFO+ room temperature probe (Double Resonance Broad Band probe, ^{19}F optimised) operating at a ^{19}F frequency of 376.5 MHz using ^{1}H decoupling, a spectral width of 111111 Hz (295 ppm), an acquisition time of 1.18 s, and 256 scans with a relaxation delay of 1.5 s. This yielded experiment times on the order of 11 minutes each and a 60 position sample changer permitted high automation (see chapter 7.1.3.1). The sample diameter was 5 mm, and the minimum volume of the NMR tube is 500 µl. The ^{19}F spectra are used to confirm purity (i.e. a single resonance) and examined for line broadening (which would indicate compound aggregation) (figure 4.3). ^{1}H decoupling also drastically reduces the complexity of the spectrum and this allows a single resonance to be defined per compound and hence makes deconvolution of hit compounds easier.

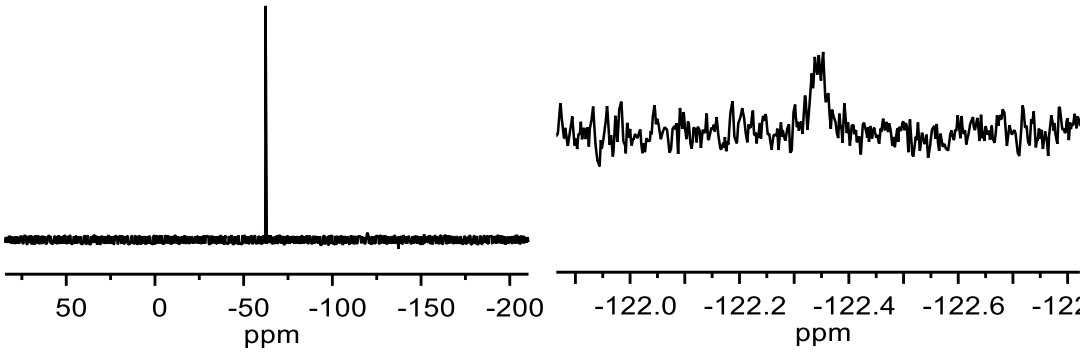


Figure 4.3: Left: 1D ^{19}F NMR spectrum of compound DDFL30. ^{19}F NMR (376 MHz, Buffer + 10% D_2O), δ -62.34 ppm. Right: 1D ^{19}F NMR spectrum of compound RAFL03. ^{19}F NMR (376 MHz, Buffer + 10% D_2O), δ -122.3. Compound RAFL03 shows severe line broadening of fluorine resonance and was not included in the final screen. See appendix 8.3.3-5 for QC analytics of all tested fragments.

Pooling fragments is essentially a method to reduce the amount of protein needed per screen and reduce the overall time for data collection. Typically, ^{1}H ligand-observed NMR methods use fewer than 5 compounds per pool.[185] However, the use of the fluorine resonance dramatically reduces spectral complexity, therefore, it is possible to screen up to 30 fluorine compounds per pool.[186] This is also dependent on the solubility of fragments and protein tolerance to organic solvents. It is necessary to design the pools in such a way as to minimise potential chemical reactions between compounds and maximise the difference between signal resonances of compounds. For example, fragments for pooling are sorted by chemical shift (lowest to highest), then the first fragment is added to pool 1, the second to pool 2, and so on, until the desired number of pools is achieved. This minimises the spectral overlap and hence,

any changes in chemical shift of fluorine resonances will be more easily identified (see figure 4.5). Another aspect to consider when pooling fragments is the presence of CF (single fluorine) or CF_3 (trifluoromethyl) containing moieties. Due to synthetic chemical methods for synthesising a fluorine containing fragment, it is usually more convenient to insert a trifluoromethyl (CF_3) group than a single fluorine atom. This is advantageous as a CF_3 group is much more sensitive and therefore a lower concentration of fragment can be used. The integral (in effect the intensity) of the resonance signal is proportional to the number of nuclei causing it, therefore a CF_3 group is 3 times more intense than a CF group (figure 4.4). This led me to sort fragments into CF and CF_3 containing pools in order to achieve a similar intensity of signals for ease of interpreting spectra. Also, many of the fluorinated compounds purchased are fluorinated aryl rings (i.e. Ph-CF_3 or Ph-F: Ph = phenyl). This meant there was a very small range of chemical shifts present in some pools. The typical range for a Ph-F chemical shift is -100 to -200 ppm, and for Ph-CF_3 is -50 to -90 ppm. Since the DDFL library consisted mainly of Ph-CF_3 containing groups, 10 compounds per pool was considered sufficient to minimise the spectral overlap. Also, 10 compounds per pool gives a final DMSO concentration of 1%, which is generally not detrimental to the protein being screened.

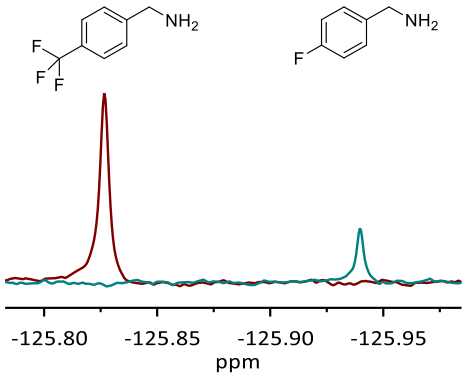


Figure 4.4: Effect of F and CF_3 on signal intensity. The CF_3 containing compound (red spectrum) is essentially 3x the intensity of the CF containing compound (blue spectrum). This influences the pooling strategy to minimise misinterpretation of spectra.

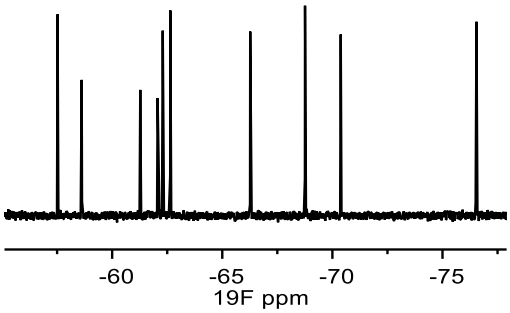


Figure 4.5: An example spectrum of a fragment pool containing 10 fragments (each present at 100 μM) in NMR buffer. Each resonance signal corresponds to a single, CF_3 containing compound. The spectrum was recorded on a Bruker AVA 400 MHz spectrometer in NMR buffer.

After QC analysis by orthogonal analytical methods (HPLC, LCMS and ^{19}F NMR), compounds were selected or removed based on the result (see appendix 8.3 for QC data). The threshold for QC analyses was > 95% purity by HPLC and a single ^{19}F resonance. Compounds that did not pass QC checks were discarded and not included in the final screen (theses are highlighted in appendix 8.3.3-5). Table 4.3 provides a summary of the final library conditions.

Library	Original number of compounds	Removal after solubility/QC analysis	Total	No. of pools	Average compounds per pool
RAFL	22	-7	**15**	3	5
CFL	24	-3	**21**	5	4.2
DDFL	63	-12	**51**	5	10.2

Table 4.3: This table shows the final library compositions after QC analysis of fragment compounds and pooling strategies. Compounds that did not pass QC checks were discarded and not included in the final screen.

This rigorous testing exemplifies how critical QC of fragments prior to use can be. This is also something not usually supplied with commercial fragment libraries. The fragment libraries were designed to use ^{19}F NMR as the primary biophysical screening method (i.e. *via* incorporation of an F atom), therefore, to develop the ^{19}F NMR screen requires optimisation of assay parameters. The practical considerations necessary are to maximise the signal and response from hit compounds. The NMR experimental parameters for fragment screening are described in the next section.

4.2.3. Developing the ^{19}F NMR Screening Parameters

To optimise the screening parameters requires a compromise between the ability to detect low-affinity binders, and the need for high throughput capacity. To validate ^{19}F NMR as an effective biophysical screening method the binding of a known fluorinated fragment to human serum albumin (HSA) was measured using a range of experimental conditions. HSA is widely used as a model system to study small molecule interactions as it binds to a wide range of ligands and is stable in many buffer solutions, therefore it is suited to a variety of assays. The medium affinity fragment (FBA, K_D = 41 μM, figure 4.6)[40] was used as a control molecule to determine optimum assay parameters for screening the fragments libraries. HSA was dissolved in the same assay buffer (20 mM HEPES, 100 mM NaCl, pH 7.5, 10% D_2O) that the fragments were measured in. A series of experiments using this system was performed to determine the optimum assay conditions for setting up the spectrometer and for gaining the maximum observable signal (see chapter 7.3.2 for materials and methods).

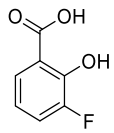


Figure 4.6: The single fluorine-containing control molecule (FBA) has a K_D of 41 μM to HSA as cited in the literature.[187] It also conforms to the fragment property filters described in chapter 3 (< 300 Da, cLogS > -4 ($[-1.67]_{calc}$), one CF, NROT < 3, H-bond D/A < 3).

^{19}F NMR ligand-observed methods involve the detection of changes in the characteristics of the ligand spectrum. In this case, changes in the 1D ^{19}F NMR spectrum will provide an indication of ligand binding, as described previously. There are several parameters that need to be optimised in order to obtain the best signal in the most efficient manner, including: ligand concentration and the number of scans in the NMR experiment. The optimum concentration for detection of a singly, fluorinated fragment (FBA) was determined to be 100 μM with 256 scans (figure 4.7). This gave a measurement time of around 11 minutes which is reasonable for a screening campaign (table 4.4). Also, the signal/noise ratio (S/N) at 256 scans, which increases with the square-root of the number of scans ($\sqrt{NS}$), is also practical for ligand detection (table 4.4).

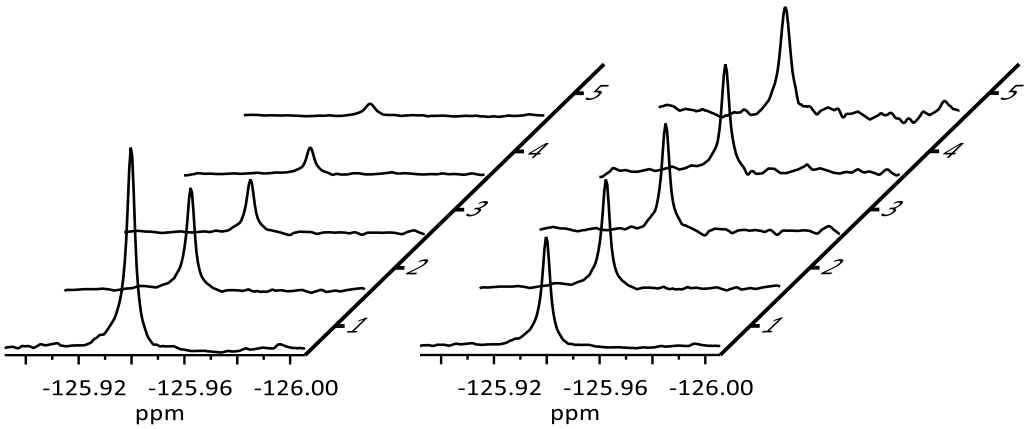


Figure 4.7: Optimisation of assay parameters. Left: The spectra show the increase in intensity of the signal on increasing the number of scans, the corresponding number of scans from 1 – 5 were 2048, 1024, 512, 256, 128. Right: The spectrum shows the effect of increasing the number of scans on the S/N ratio (spectra are intensity normalised).

No.	No. of Scans	Intensity Ratio	S/N Ratio	Experiment time
1	2048	1.00	131.26	01:31:28
2	1024	0.51	94.25	45:44
3	512	0.27	70.56	22:54
4	256	0.13	49.46	11:25
5	128	0.06	32.09	5:42

Table 4.4: Overview of the NS, S/N and experiment time. Calculated intensity ratios and signal/noise (S/N) ratios of a singly fluorinated compound (FBA) as a function of number of scans and the experiment time (see chapter 7.3.2 for materials and methods). The spectra were analysed with MestReNova NMR suite.

4.2.4. Protein:Ligand Concentration Ratio

A concentration series was measured for the FBA compound against HSA to determine the optimum protein:ligand ratio for the maximum response. FBA was shown to bind to HSA by a distinct line broadening of the signal in the bound state. The K_D of FBA is 41 μM which is approximately in the intermediate chemical exchange rate range (~k_{off} = 410 s^{-1}), therefore line broadening is the predominant observed effect. This experiment shows the sensitivity of the ^{19}F assay to the binding event (see figure 4.8).

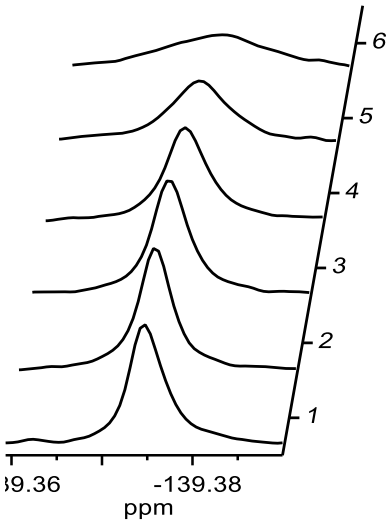


Figure 4.8: Line width broadening of control molecule FBA (K_D = 41 μM). The plot shows the effect of increasing the HSA concentration on the fluorine linewidth. The compound concentration was 100 μM and the HSA concentration ranged from: 1) blank; 2) 100 nM; 3) 500 nM; 4) 1 μM; 5) 2.5 μM; 6) 5 μM.

It is clear that increasing the concentration of HSA increases the fraction of the bound ligand. Even with a small fraction of the bound ligand (table 4.5, spectrum 5, p_b = 0.018) the sensitivity of the fluorine nucleus to changes in it's electronic environment show are observable by an increase in linewidth and consequent decrease in intensity. Table 4.5 shows the absolute values for intensity and line width changes for increasing the HSA concentration.

Spectrum	HSA concentration (μM)	Fraction Ligand Bound ($x10^{-3}$)	Intensity ($x10^6$)	Intensity Ratio	Line width (Hz)
1	0	0	5.07	1	1.57
2	0.1	0.7	4.97	0.98	1.75
3	0.5	3.5	4.79	0.94	1.81
4	1	7.1	3.98	0.78	2.02
5	2.5	17.6	2.67	0.53	3.10
6	5	35.1	1.45	0.29	5.48

Table 4.5: The effect of line broadening and intensity changes on increasing the bound fraction of FBA. FBA was kept at a constant concentration (100 μM) and the protein (HSA) concentration was varied. The intensity ratio is defined as the ratio between the intensity of the protein containing sample and the non-protein containing sample.

To gain the maximum signal after a binding event, is therefore a compromise between ligand and protein concentration, i.e. the maximum signal gained is with the largest fraction of ligand

bound (table 4.5). The volume of the NMR tube (5 mm sample diameter) requires 500 µl of solution, this influences the amount of protein needed to gain a sufficient ligand:protein concentration ratio whilst also limiting protein consumption. The maximum fraction of the ligand bound to the protein will give the greatest signal, however, we need to maintain an observable signal using the minimum amount of protein. For singly fluorinated compounds (CF), 100 µM is the minimum concentration for detection by ^{19}F NMR (as stated by Bruker). From the analysis of FBA binding to HSA (table 4.5), a significant line-broadening effect is observed at $p_b[L] = 0.018$ (figure 4.8, spectrum 5 and figure 4.9, yellow dashed line) and the signal intensity reduces by approximately 50%. This fraction of ligand bound ($p_b[L] = 0.018$) is therefore the limit to reliably observe a ligand binding event. To determine the optimal ligand/protein ratio, a simulation of the fraction of ligand bound (at constant ligand concentration of 100 µM) vs K_D at varying protein concentrations was performed (figure 4.9).

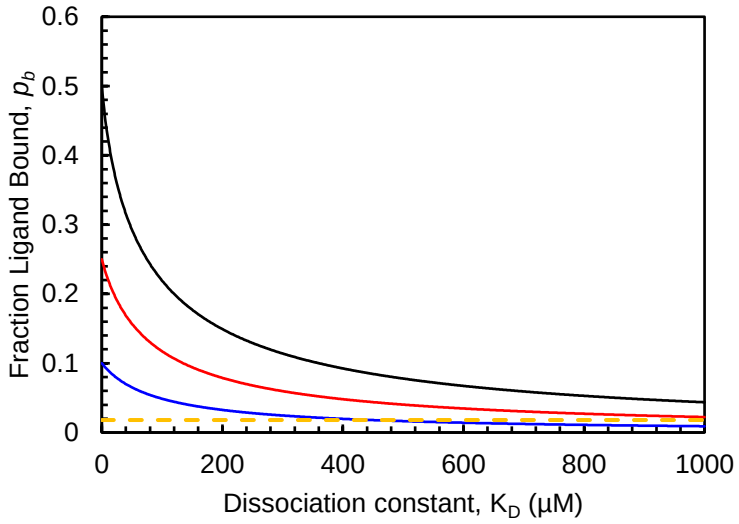


Figure 4.9: Simulation of fraction of ligand bound vs dissociation constant, K_D. The plot shows a simulation of the fraction of ligand bound vs the dissociation constant at different protein concentrations (ligand concentration is constant at 100 µM). The black curve is $[P]_0 = 50$ µM, red curve is $[P]_0 = 25$ µM and the blue curve is $[P]_0 = 10$ µM. The yellow dashed line indicates the threshold value (0.018) for reliable detection of the bound species under these experimental conditions.

A simulation using a constant protein concentration of 10 µM (blue curve, figure 4.9), shows an intersection at approximately 500 µM. This therefore implies that, under these conditions, we could reliably see compounds in the range of 500 µM K_D and lower which is very good for a fragment-based screen. A 10 µM protein concentration equates to 5 nanomoles of material (500 µl sample volume), therefore for small/medium sized proteins (10 – 30 kDa) this equates

to 50 – 150 μg of protein per pool. There are 13 pools covering all fragments in the library, this gives a range for amount of protein as 0.65 mg to 1.95 mg (10 – 30 kDa) per screen. Typically, the yield of protein from a 1 L bacterial expression culture is 1 – 10 mg which is sufficient for multiple screens and fragment hit validation. The final conditions for screening were determined to be $[L]_0$ = 100 μM and $[P]_0$ = 10 μM. This would give a practical sensitivity for the fluorine signal, a reasonable throughput for protein consumption and a quantifiable measure of the fraction of ligand bound.

4.2.5. ^{19}F NMR Screening of the Ubiquitin Conjugating Enzyme UbE2L3

As described in chapter 2, the ubiquitin conjugating enzyme E2L3 (UbE2L3) is often referred to as "difficult" or "undruggable" target. Due to this target status, we expected that all possible chemical and biological matter might be needed to find access points into further drug discovery. In this work we applied the newly developed OPPG technique (see chapter 2.3) which, in a first round using the pF-Phe centred dipeptide library as starting point, turned out to only deliver very low affinity binders and in addition was not showing increased affinity with library length expansion. While a variation of fluorinated, non-natural amino acid centred dipeptide libraries might support the identification of new targetable binding pockets, we progress in this work with the screening of our in-house designed ^{19}F-fragment libraries, see Chapter 2. The in-house, designed, fluorinated fragment libraries (Chapter 3) were evaluated against this protein to discover novel ligands. The protein was expressed and purified (described in Chapter 2.3.3) and dialysed into NMR assay buffer (excluding D_2O: 20 mM HEPES, 100 mM NaCl, pH 7.5) (see chapter 7.2.3 for materials and methods).

4.2.5.1. Enzymatic Functional test of UbE2L3 (Carried out with Dr Joanna Koszela)

The functional viability of the UbE2L3 enzyme was assessed to confirm the protein is active under the same conditions as applied for the NMR screen (i.e. buffer, temperature, pH and DMSO concentration). A simple, gel-based enzymatic assay was used to measure the activity of E2 conjugating enzymes. This involves incubating a mixture of the E2 enzyme, an E1 enzyme, ubiquitin protein and ATP (see figure 4.10). The E2 protein is verified as active by a corresponding increase in molecular weight on a SDS-PAGE gel (figure 4.11).

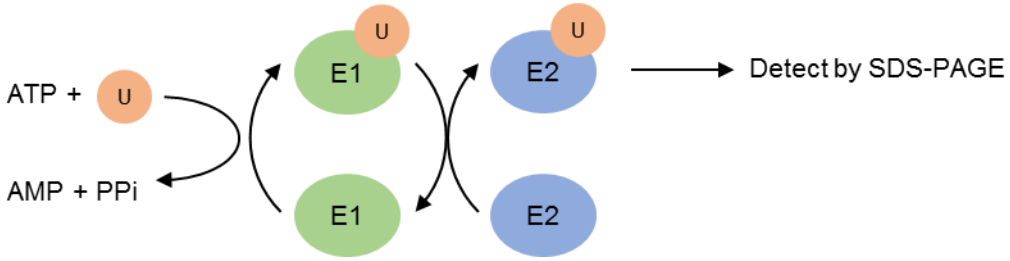


Figure 4.10: Schematic showing the initial steps of the ubiquitin proteasome cascade. The E2 enzyme is 'loaded' with ubiquitin via a covalent thioester bond, and the change in molecular weight is detected by SDS-PAGE analysis.

The fortunate characteristic of UbE2L3 is that it forms a covalent complex with only a single ubiquitin protein. Therefore, a single band at ~ 27 kDa (UbE2L3 + ubiquitin) is observed (L4, figure 4.11) corresponding to the E2-ubiquitin complex.

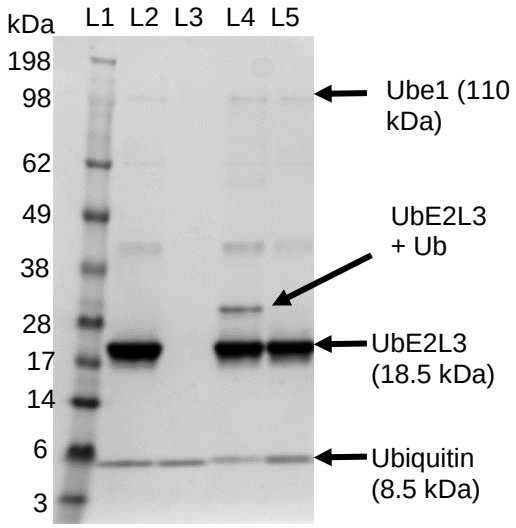


Figure 4.11: SDS-PAGE gel for the UbE2L3 functional enzymatic assay. L1: molecular weight marker; L2: no ATP control; L3: no UbE2L3 control; L4: ubiquitin, ATP, E1 and UbE2L3; L5: DTT containing control (see chapter 7 for materials and methods).

4.2.5.2. ^{19}F NMR Screening of the UbE2L3 enzyme

All 13 fragment pools from the 3 CTB libraries (RAFL, CFL, DDFL) were analysed for chemical shift changes and/or line broadening effects on the addition of 10 μM UbE2L3. However, there were no evident perturbations to the chemical shift on addition of protein. This indicates that there were no fragment hits in the fluorine fragment libraries. The 1D ^{19}F NMR spectrum shown in figure 4.12 is an example superposition from the DDFL library containing 11 fragments. The blue spectrum is compound only and the red spectrum contains 10 μM UbE2L3. There is no significant change in peak shape, intensity or chemical shift of any

fragments in this pool. Therefore, it is considered that there are no binding interactions to the protein occurring.

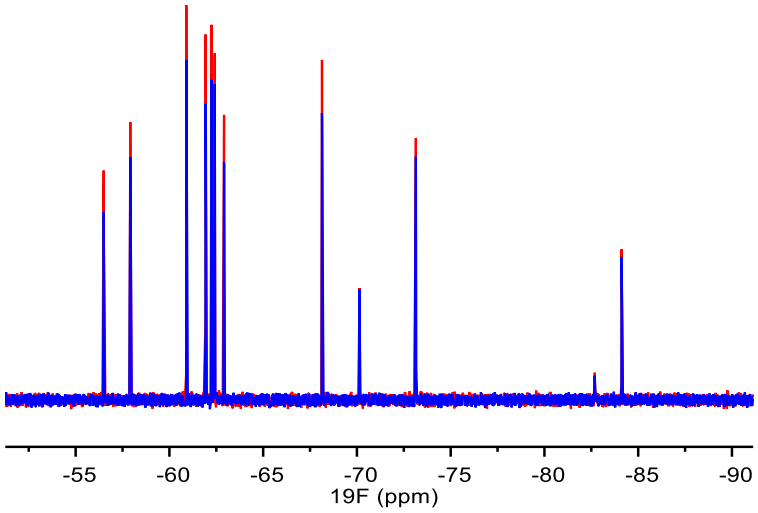
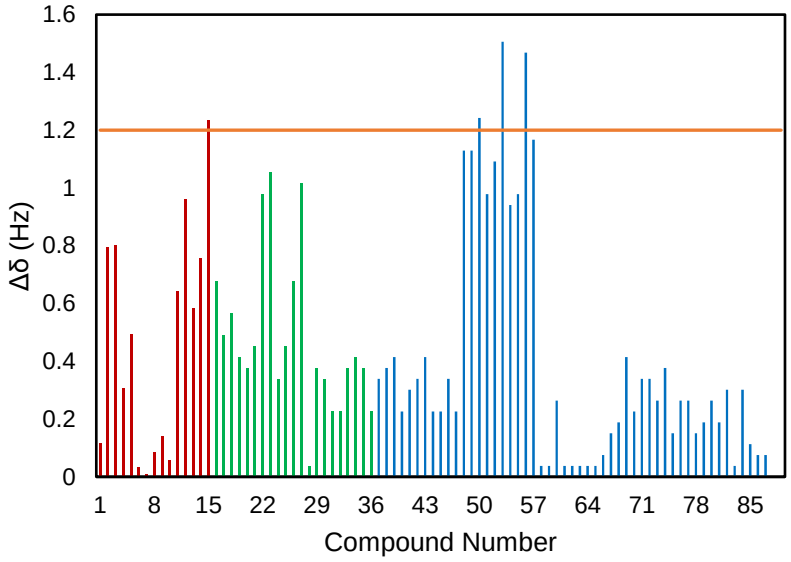
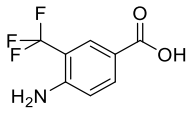
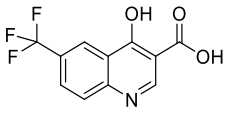


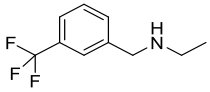
Figure 4.12: Superposition of no protein (red) and protein containing (blue) 1D ^{19}F NMR spectra. It is an example pool of fragments from the DDFL library (11 fragments in the pool). There is a small decrease in intensity of some peaks on addition of protein, however, this was not considered as strong enough to induce a start of chemistry efforts for identification of promising hit compounds.

Figure 4.13 shows a histogram of the observed chemical shift change (Δδ) for all compounds screened against UbE2L3. Using a threshold of 3 standard deviations from the baseline, the calculated threshold value is 1.2 Hz. This identifies 4 compounds with a chemical shift above this value (figure 4.13). The largest change in chemical shift is Δδ 1.50 Hz (deconvoluted to compound DDFL09). However, such a small change in chemical shift indicates a low affinity binder (> 1 mM) and it was therefore considered that considerable chemical effort would be needed to increase the Ube2L3 binding affinity of all the primary hit fragments. Also, gains in affinity resulting from SAR analysis would likely be minimal as each compound is structurally dissimilar. Therefore these fragments were not pursued further. This would also have prevented to follow up with the main goal of the thesis, namely to design, establish and explore new ^{19}F-fragment libraries in the Auer lab. It was therefore decided to focus efforts towards a next PPI target of different origin and thereby experiment with possible binding pockets which might be hit by our libraries.









DDFL09	**DDFL39**	**DDFL53**	**RAFL14**
Δδ 1.50 Hz	Δδ 1.47 Hz	Δδ 1.24 Hz	Δδ 1.17 Hz

Figure 4.13: Top: Histogram of ^{19}F NMR chemical shifts of each compound screened against UbE2L3. The compound concentration was 100 μM and protein concentration was 10 μM. The change in chemical shift (Δδ) was calculated by subtracting the resonance frequencies of the protein containing spectrum from the compound only spectrum. The solid orange line indicates a threshold value that is 3 standard deviations from the Δδ Hz. This was used as the cutoff for identification of possible ligands. The red bars are the RAFL library compounds; green: CFL library; blue: DDFL library.

Target druggability is a term used in drug discovery to measure the likelihood that a molecular target will bind with high affinity to a small, drug-like molecule. A high druggability score is usually assigned based on the hit rates from HTS or FBDD campaigns, for example, Hajduk *et al* refer to fragment screens as a prediction of the success of FBDD.[188] The concept has gained attention in Pharma where a high probability for success is paramount, for example, AstraZeneca are now using target druggability as a measure on which targets to pursue.[189]

To continue pursuit of this target may require extension to the chemical space coverage of the fluorinated fragment libraries. Using the method described previously from Dalvit *et* al, the

3-FABS NMR screening approach could be applied here. This is an enzymatic reaction in which ubiquitin is covalently attached, via a thioester, to the E2 conjugating enzyme. Using a fluorinated amino acid at the N-terminus of ubiquitin would provide a useful probe for monitoring this reaction. Screening with non-fluorinated fragments (hence a greater chemical space coverage) would then be a possibility to discover novel ligands. Also, many fragment screening approaches based on ^{19}F utilise a T_2 relaxation based experiment as this parameter can be more sensitive.[190] This is a more direct measure of binding and is discussed in more detail in chapter 5.

4.3. Concluding Remarks

In conclusion, the designed fluorine fragment libraries (RAFL, CFL and DDFL, chapter 3) were established for use in a ^{19}F NMR-based screening approach, then quality checked by multiple, orthogonal biochemical methods (HPLC, LCMS and ^{19}F NMR) to determine purity and identity. Compounds that did not pass these QC checks were discarded for use in the final screen. Compounds were also sorted by ^{19}F chemical shifts and designated to specific pools to minimise spectral overlap. The optimum concentrations for ^{19}F NMR screening parameters were a $[L]_0$ of 100 μM and $[P]_0$ of 10 μM. This gave a final molar ratio of 10:1 (ligand:protein) to provide a reasonable throughput for protein consumption and a visible binding response. A consideration for future screening conditions is to use a lower concentration of CF_3 containing fragments. The DDFL compounds all contain a CF_3 moiety, therefore for future screening campaigns the fragment concentration can be reduced to maximise the complex concentration and, hence the fluorine signal response.

It is also shown how important that rigorous QC of fragments be performed prior to screening. All compounds are checked for aqueous solubility, aggregation and purity to give the highest quality fragment libraries and fragments that did not pass QC checks were discarded.

UbE2L3 was screened for hit fragments by ^{19}F NMR, however this fragment screen yielded hits with high likelihood of mM affinity. Chemical expansion was considered too demanding to tackle in this thesis, and gains from SAR analysis were not thought to present a worthwhile advantage given the limited capacity. Screening of other PPIs is therefore necessary to assess the diversity and deliverability of the fragment libraries. Also, screening multiple targets with the same library can assess the quality of diversity and fragments. It is highly likely that screening multiple PPIs would not deliver the same success.

In summary, the collection and analysis of fluorinated fragments was carried out and an initial screen with UbE2L3 was performed. It was decided to exploit and analyse the fragment libraries against another PPI target to assess the qualities of the libraries.

Chapter 5

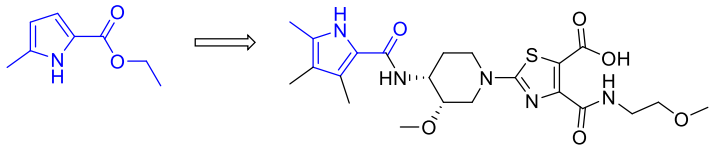
Fragment Based Screening of the Bacterial PPI NusE/NusB

The work presented in this chapter describes the screening of the bacterial PPI target NusE/NusB by ^{19}F NMR with the fluorinated fragment libraries described in Chapter 3. The introduction presents the target and elucidates the importance of this PPI in the context of a potential antibiotic development. In the results section the details of the fragment screen and follow up of confirmed hit compounds are presented. The chapter concludes with describing the future scope of experiments and personal perspective.

5.1. Introduction

5.1.1. Antibiotics

Due to antibiotic resistance, many bacterial infections have become impossible to treat,[191] particularly multidrug resistant strains for example the ESKAPE pathogens: *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa* and *Enterobacter*. Therefore, the identification of new targets and resistance pathways is essential for developing next-generation antibiotics to combat this. An example of a recent success story that aims to target this issue is the identification of AZD5099 by AstraZeneca. AZD5099 was identified in a fragment-based screen that targeted the bacterial topoisomerase II enzyme.[192] Multiple rounds of optimisation to compound AZD5099 resulted in it entering clinical trials for the treatment of bacterial infections caused by Gram-positive and some Gram-negative bacteria. As shown in figure 5.1, a pyrrolamide scaffold identified in the fragment-based screen was developed into a potent (< 100 nM IC_{50}) inhibitor of bacterial topoisomerases.



5.1 Fragment Hit **5.2** Final clinical candidate

Figure 5.1: Initial fragment hit (**5.1**) against the bacterial topoisomerase II enzyme (left) and the final clinical candidate (**5.2**, AZD5099, right) that was developed through many rounds of medicinal chemistry and is currently in Phase I clinical trials.

However, very few new drugs are in clinical development to treat bacterial resistant infections. As described previously, PPIs are an emerging class of drug target and there are now a large number of publications showing that small molecule inhibition of PPIs is a feasible approach.[193,194] A mounting body of research is now investigating the bacterial interactome to identify essential PPIs involved in bacterial growth and resistance.[195]

5.1.2. Introducing the Target – NusE/NusB Protein-Protein Interaction

PPIs involved in transcription are an attractive option for pharmacological intervention.[196] A host of regulatory proteins are involved in this complex pathway, and disruption of this process has been shown to be of therapeutic benefit.[197,198] The transcription of the genome during RNA synthesis is a complex cycle consisting of three major stages: initiation,

elongation and termination. The enzyme at the core of this process is RNA polymerase (RNAP) and it is regulated by a host of N-utilisation substance (Nus) transcription factors (NusA, NusB, NusE and NusG). These factors interact directly and/or indirectly with the RNAP to moderate its function and to produce a stable elongation complex (EC).

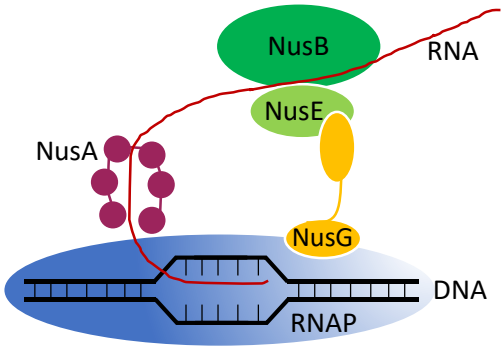


Figure 5.2: Schematic of the roles of the Nus factors in the transcription elongation complex. The NusE/NusB heterodimer binds ssRNA and is anchored to the RNAP through NusG interaction. NusA is a multidomain protein that binds close to the RNA exit channel on the RNAP. Figure adapted from Drögemüller.[199]

NusA is one of the first proteins to bind the RNAP after initiation and binds near the RNA exit channel, where it most likely prevents reattachment of the αCTD of the RNAP to DNA, thereby preventing a stalled transcription complex.[200] The NusG protein is a transcription regulator found across all prokaryotes, eukaryotes and archaea.[200] It has a two domain structure connected via a flexible linker that infers it is able to interact with multiple proteins as a linker.

NusE (also called ribosomal protein S10) is a small, 12 kDa, protein and, as described, forms a separate complex with NusB that is also involved in the antitermination process of prokaryotic cells. In addition to its role in transcription, NusE is involved in linking transcription and translation as a component of the 30S subunit of the ribosome. The 30S subunit is the smaller subunit of the 70S ribosome; it complexes with the larger 50S subunit to form the 70S prokaryotic ribosome. The 30S subunit is further formed of the 16S ribosomal RNA (rRNA) and 22 proteins. The 30S subunit is the protein target for tetracycline and the aminoglycoside class of antibiotics.

NusB is also a small, 16 kDa, protein involved in bacterial transcription of rRNA. It is recruited to bind to the *boxA* RNA sequence and its affinity for the *boxA* sequence is greatly increased in the presence of NusE. NusB is also involved in modulating the transcriptional elongation rates.[201]

Antitermination is unique to prokaryotes and is the cell's aid to fix premature termination of RNA synthesis during transcription is an interesting target for drug intervention. Transcription and translation in prokaryotes are directly coupled and transcription antitermination allows the cell to progress from early gene expression to delayed early gene expression. In the complete antitermination complex, the Nus factors interacts with the RNA polymerase (RNAP) and allows progressive transcription elongation through termination sites. This is essential for efficient transcription of bacterial ribosomal RNA operons.[202,203–205]

One particular PPI of this complex governing the regulation of transcription is the interaction between transcription factors NusE and NusB (collectively known as NusE/NusB. This PPI is critical for the formation of the highly regulatory antitermination complex in prokaryotes and to allow stable RNA transcription. The importance of the NusE/NusB binding interaction was described separately by Robledo (1991), Court (1995) and later by Lou (2008).[206–208] It was shown that by incorporating point mutations separately into NusE and NusB resulted in a reduced protein-protein binding affinity, which subsequently affected the formation of the antitermination complex. Furthermore, the mutations affected the ability of *E. coli* cells to efficiently transcribe the 16S and 23S ribosomal transcripts, therefore leading to a reduced number of new ribosomes and, hence, a decrease in cell growth.

5.1.2.1. Structure of NusE/NusB interface

The PPI surface area of the NusE/NusB complex is determined to be around 1600 Å^2 and comprises a mixture of hydrophobic and hydrophilic interactions, arising from the α1-helix and β2-strand of NusE facing the helical bundles of NusB (figure 5.3).[209] The key interactions have been studied extensively by X-ray crystallography and NMR spectroscopy and there are currently 7 structures of this complex (from various organisms) available in the Protein Data Bank (PDB); the NusEΔ is also referred to as S10Δ^{loop} or NusEΔ^{loop} in scientific journals.

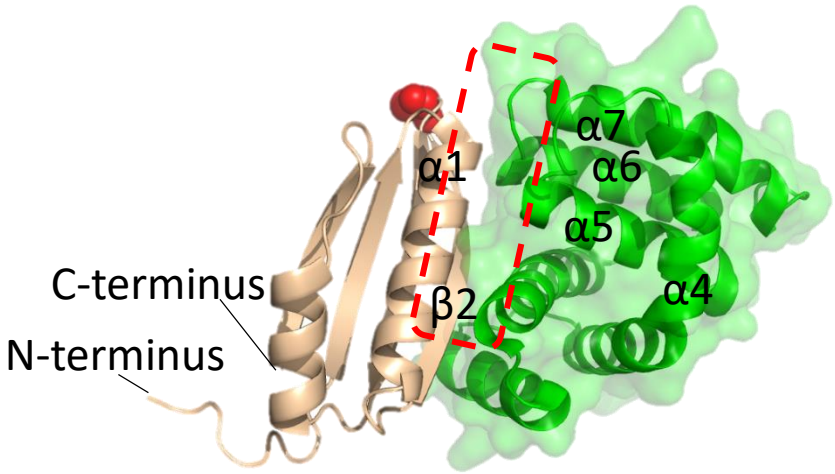


Figure 5.3: Crystal structure of the NusEΔ/NusB protein complex (PDB ID: 3D3B) from *E. coli*. NusEΔ is coloured pale, and NusB is coloured green. The red dashed box indicates the interface of the main protein-protein interaction between the helical bundle (α4-7 of NusB) and the α1 helix/β2 strand of NusE. The red spheres on NusEΔ denotes Ser46, which replaces the ribosome binding loop region (46-67) in this construct. This image was constructed in PyMol and adapted from Cossar *et al*.[209]

The key electrostatic interactions derived from molecular structures are shown to be conserved across many, clinically relevant bacterial strains, including: *Staphylococcus aureus*, *Streptococcus pneumoniae* and *Haemophilus influenza*.[210]

5.1.2.2. Inhibitors of NusE/NusB PPI

At the start of this project, there were no known inhibitors of this bacterial PPI, however, in July 2017 a group (Cossar *et al*, University of Newcastle, Australia) published a series of compounds targeting the NusE/NusB PPI.[209] To support the hypothesis, Cossar *et al* demonstrated the ability to inhibit the NusE/NusB PPI by screening a 9-mer peptide (HO-YDHRLLDQS-NH_2), in a bacterial growth assay. This peptide replicates the α1-helix of NusE and returned an IC_{50} of 71 μM, which confirms the potential to inhibit this PPI. They then used a pharmacophore-based screening approach of the mini-Maybridge 56,000 compound library that yielded 25 preliminary hits, which was reduced to 5 after energy minimisations (figure 5.4). This compound library was used in a classical HTS screening approach, as compounds are larger and more drug-like, which explains the poor hit rate (0.04%). Compounds were re-synthesised and tested in a bacterial growth inhibition assay and 3 compounds inhibited the NusE-NusB binding interaction with > 50% efficiency.

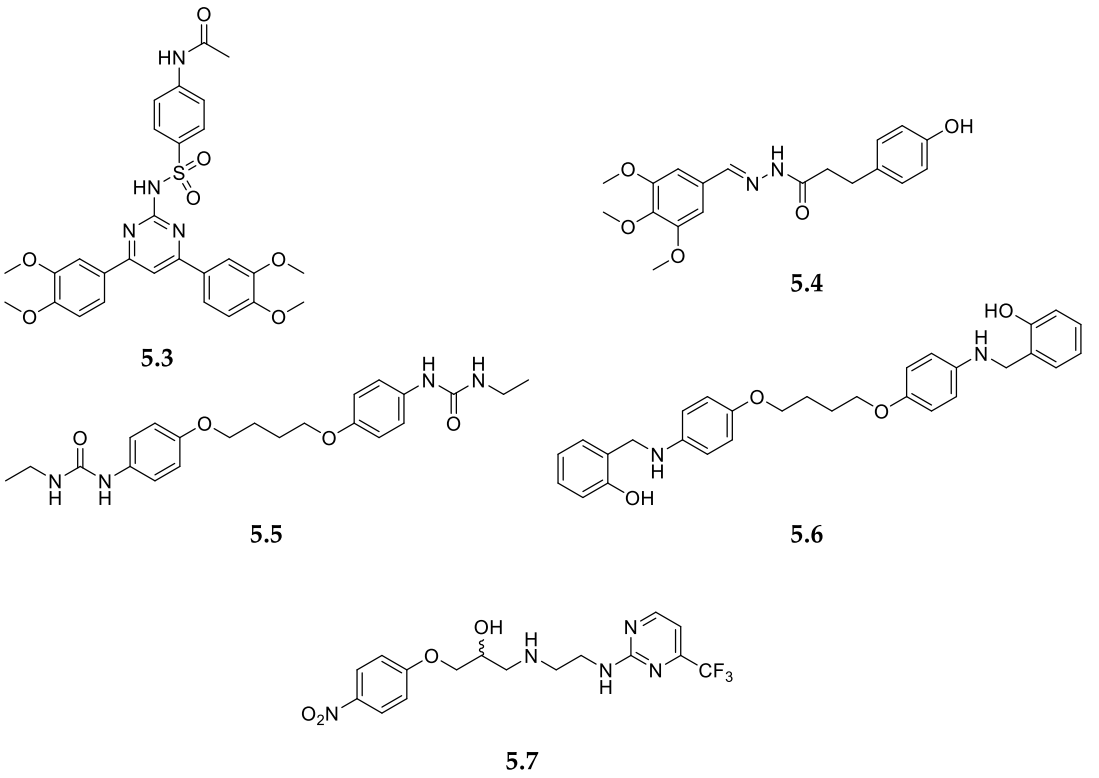


Figure 5.4: Chemical structures of compound hits identified from a pharmacophore-based screen of the Maybridge 56,000 compound library. Figure adapted from Cossar *et al*. [210]

These results validate the NusE/NusB PPI as a new and accessible target for pharmacological intervention. A fragment-based approach may provide a better starting point for chemical tractability and discovery of novel starting points.

5.1.3. Aims of this work

The aim of the work described in this chapter was to discover novel fragment hits against the NusE/NusB bacterial PPI. This PPI complex was screened against the fluorinated fragment libraries (Chapter 3) by ^{19}F NMR chemical shift perturbation (CSP) analysis (see chapter 4). The protein was supplied by collaborators in the Department of Biopolymers at the University of Bayreuth, Germany, and prospective follow-up of hits was feasible using a 'SAR by catalogue' approach (see chapter 3).. Furthermore, access to more specific NMR experiments and expertise with the Biopolymer group allowed for further development and evaluation of fragment hits.

5.2. Results and Discussion

The NusEΔ/NusB protein complex was supplied by collaborators at the University of Bayreuth, Germany. The delta, Δ, refers to a deletion of a 22 amino acid ribosome binding sequence of NusE to increase the stability of the overall complex. It was found that the loop region does not contribute to the overall transcription efficiency so does not impact on complex stability.[208] Furthermore, in this format it is unknown what the K_D of the complex is. The data referring to protein handling is described in the materials and methods section, chapter 7.4.

5.2.1. Fragment Library Screening by ^{19}F NMR of NusEΔ/NusB PPI

The NusEΔ/NusB protein complex was screened in a ^{19}F NMR assay on a Bruker Avance 400 MHz spectrometer against the three designed fluorinated fragment libraries (RAFL, CFL and DDFL, see chapter 3). Using ^{19}F NMR, the binding of a ligand to a component of the protein complex is detected, not necessarily the disruption or disassociation of the complex in solution. Fluorinated fragments were present at 100 μM and NusEΔ/NusB was added at 10 μM for protein containing samples (see chapter 7.4.2 for materials and methods).

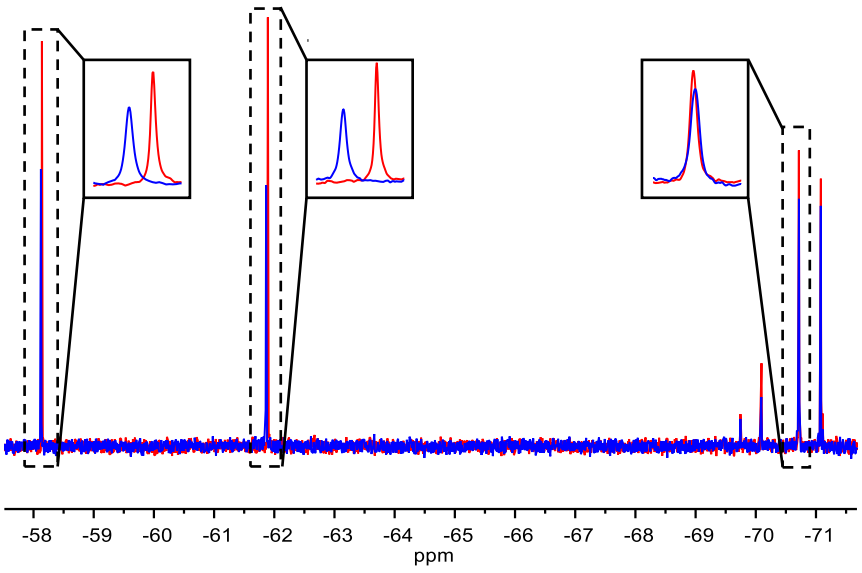


Figure 5.5: Superposition of two 1D ^{19}F NMR spectra from a compound pool containing 4 fragments. The red spectrum is compound only, the blue spectrum is recorded in the presence of 10 μM NusEΔ/NusB. The two peaks at approximately -58 ppm and -62 ppm are significantly downfield shifted in the presence of protein, indicating a binding event. The peak at – 70.6 ppm shows no change in chemical shift on the addition of protein and is shown to demonstrate a non-binding fragment.

Figure 5.5 shows a superposition of the compound only spectrum (blue) and protein containing spectrum (red) of the fragment pool containing hits. There are 4 compounds in this pool and two of the compounds show a change in the chemical shift on addition of 10 μM NusEΔ/NusB. This indicates a change in the chemical environment of the fluorine atom and hence a binding event. Deconvolution of hit fragments is simple due to the single ^{19}F NMR chemical shift signal. The compounds were part of the 'Clustered Fragment Library' collection, which therefore could be argued that this is the best method for library design. However, with such a small number of compounds, the diversity of each library is likely similar, therefore identifying compounds from the CFL library is most likely fortunate. From herein, compound hits are named by the library they originated from, e.g. CFL, and the ID within that library. The two initial fragment hits are thus named CFL02 and CFL17 (see figure 5.7). Using the ^{19}F NMR conditions described previously (see Chapter 4), figure 5.6 shows a bar chart of the observed chemical shift (Δδ in Hz) change for each compound contained within all libraries (87 compounds in total). The threshold for hit identification was set to 4 Hz (0.01 ppm). The entire screen was completed in around one week and two compounds were identified as hits and selected for follow up.

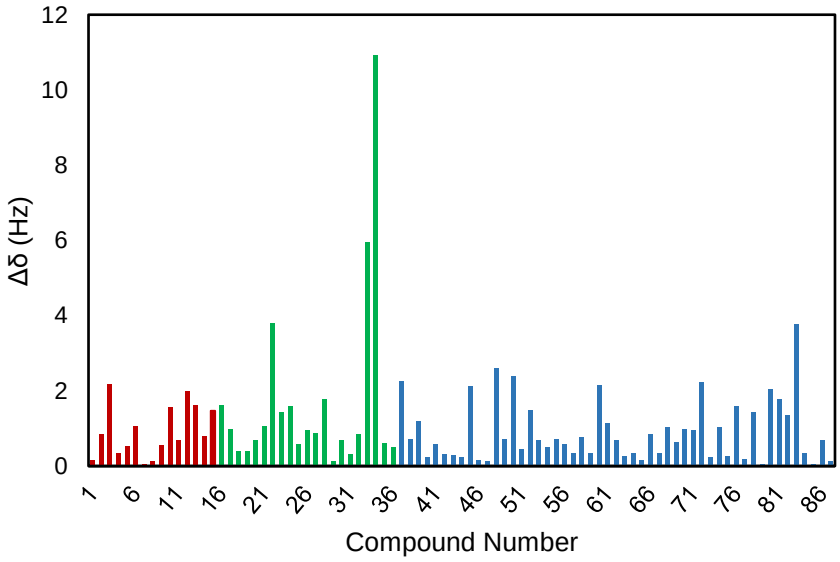
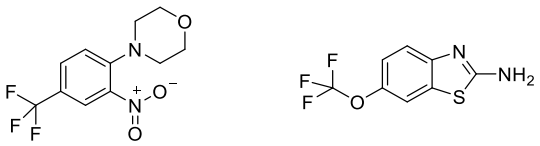


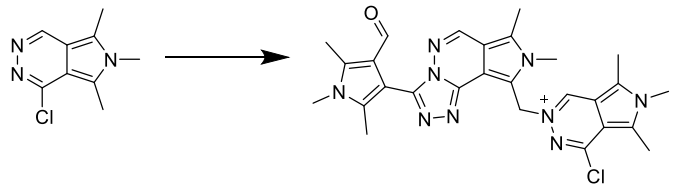
Figure 5.6: Change in chemical shift for all fluorinated fragments screened vs the NusE/NusB PPI protein target. The red section is the RAFL library fragments, green is the CFL library fragments and blue the DDFL library fragments.



CFL02 CFL17

Figure 5.7: Chemical structures of fragment hits identified from ^{19}F NMR screen of NusEΔ/NusB

The most productive step after hit identification and deconvolution is to check that the binding is not a false positive. These can arise due to aggregation of compounds in solution, which is particularly problematic in fragment-based screening where high concentrations of ligand are used (up to 1 mM). However, there are other mechanisms giving rise to false positives, for example, Walden *et al* describe the analysis of a promising hit fragment to the UbE2T protein only to discover a contaminating zinc metal ion was solely responsible for the observed binding effect.[211] Concerns arose after analysis of hit analogues appeared to be inactive and a co-crystal structure confirmed the binding of the metal to the active site cysteine. Another example of a false positive screening hit comes from Klebe *et al* who discovered a fragment hit (figure 5.8, **5.8**) against the aspartic protease endothiapepsin. The compound was a hit in five out of six fragment screens including: TSA, biochemical assays and ITC.[212] However, again, X-ray crystallography revealed that the fragment underwent a self-reaction mechanism and rearrangement in solution to produce a larger isoform of the original hit (**5.9**, figure 5.8).



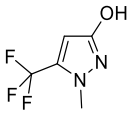
5.8 Hit Fragment **5.9** Rearranged isoform

Figure 5.8: Fragment rearrangement from screen of aspartic protease endothiapepsin. Reproduced from Klebe *et al*.[212]

In both of these papers, X-ray crystallography is used as an orthogonal binding confirmation method, and validation of the hit fragments. These are two excellent examples of using separate biophysical detection methods for validating hit fragments. Without these rigorous tests for artefacts could result in pursuing an erroneous compound costing precious time and money. The next stage therefore was to validate the hits to determine that they are not artefacts and to obtain biophysical binding data, such as dissociation constants.

5.2.2. Fragment Hit Validation

The ^{19}F nucleus spans a large chemical shift range, *ca.* 800 ppm, and is very sensitive to changes in the local electronic environment. These changes can occur due to ligand binding, or minute differences in pH and DMSO concentration. To confirm the binding of fragments, and to determine that shift changes were not due to, changes in experimental conditions hit compounds were analysed individually by ^{19}F CSP analysis (figure 5.10). The chemical shift changes of hit fragments was reproducible, when tested individually, giving confidence that these are real hits. Also, a second control compound, to reproduce the binding effect (figure 5.9) was included. This compound did not show any binding in the initial fragment screen (figure 5.11) therefore it was used as a control molecule to reference the change in chemical shift of the hit fragment.



CFL23

Figure 5.9: A control molecule that did not show binding in the initial fragment screen.

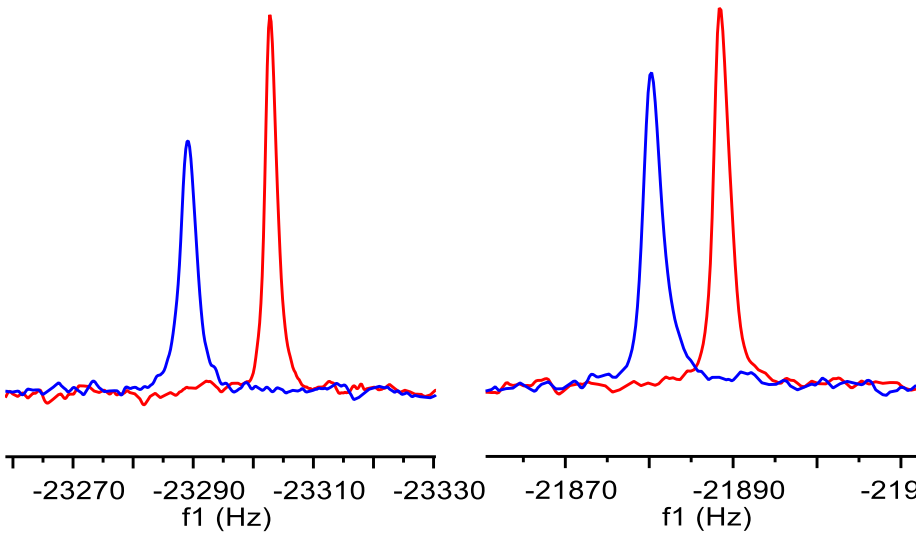


Figure 5.10: Superposition of 1D ^{19}F NMR spectra. Changes in ^{19}F chemical shift for the two hit fragments are demonstrated in the presence (blue spectra) and absence (red spectra) of 10 μM NusEΔ/NusB Left: compound CFL02 and right: CFL17.

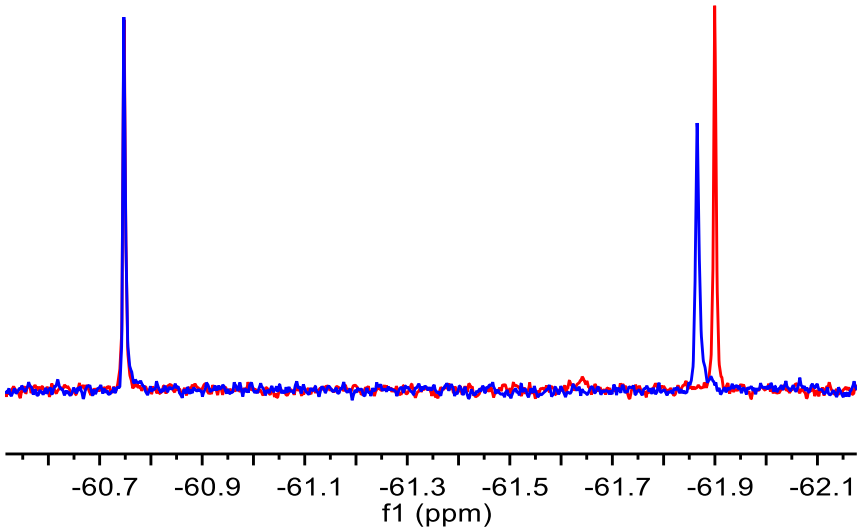


Figure 5.11: A control molecule (figure 5.2.3) at -60.75 ppm was included in a 1:1 mixture with the fragment compound (100 μM) plus NusEΔ/NusB to show that the CSP of the CFL02 fragment is reproducible. The control molecule is used as a reference to show the change in chemical shift of the hit fragment. Transverse Relaxation Experiments

To validate the binding effect, and to recognise that the ^{19}F chemical shift perturbation (CSP) represents true complex formation, a second NMR-based method was used. In this instance, validation of NusEΔ/NusB binding is necessary due to the sensitivity of CSP to changes in DMSO concentration and pH. Therefore, this hit validation experiment was performed to clarify any possibility of false positive interactions occurring. The resonance signals of molecules in ligand-observed NMR have been used to study protein-ligand complexes by exploiting differences in their relaxation and diffusion rates. Here, I describe the use of a ^{19}F relaxation experiment to monitor ligand binding.

5.2.2.1. $T_1\rho$ relaxation detected ligand binding - Theory

$T_1\rho$ is the spin-lattice relaxation in the rotating frame and is the mechanism by which the magnetisation vector decays along the applied radiofrequency field, which is static in the rotating frame of reference. With this method one monitors the longitudinal relaxation (analogous to T_1 decay constant) in the rotating frame by application of a spin-lock pulse sequence to yield a signal intensity of:

$$I_0 \exp\left(\frac{-\tau}{T_1\rho}\right)$$

Equation 5.1

Where I_0 is the intensity, $T_1\rho$ is the decay constant in the rotating frame and τ is the spin-lock pulse length. The experiment is repeated with different values of τ and the resulting intensities used to find the value of $T_1\rho$. This detection method is used to exploit the difference in the relaxation rate between the bound and free states of the fragment molecules. The ^{19}F $T_1\rho$ filter is achieved with a composite pulse scheme applied before acquisition that attenuates signals of rapidly relaxing ligands. A fast tumbling molecule, for example a small molecule ligand, will experience a slow relaxation and gives rise to sharp peaks. A large, slow tumbling molecule, e.g. a protein macromolecule will experience a fast relaxation and gives rise to broad peaks. Any bound ligands will adopt the properties of the protein therefore binders show strong attenuation of the signal, i.e. $T_1\rho$ is directly dependent to the overall molecular rotational correlation time. The basic pulse sequence is-based on incorporation of a spin-echo period. Initially, a 90° pulse creates the transverse magnetisation; a train of 180° pulses determines the decay of the magnetisation and is repeated *n* times (L4, figure 5.12); acquisition is then performed as usual. The delay between pulses (D_{20}) is a fixed period and the intensity of the signals will decrease when the spin-lock period is increased.

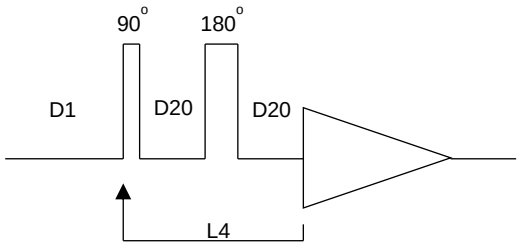


Figure 5.12: A schematic of the $T_1\rho$ composite pulse sequence. D_{20} is the delay between 180° pulses (wide rectangle) and L4 is the number of repetitions. By definition the spin-lock (SL) time (in milliseconds) is therefore: $SL(ms) = L4 \times (2 \times D_{20} + p2)$. D_{20} is a fixed echo time, set to 10 ms, to allow elimination of diffusion and J-modulation effects. The triangle represents the FID for detection.

5.2.2.2. Relaxation-detection of ligand binding

Using the $T_1\rho$ approach described in section 5.2.2.1, fragment hits were validated by this method. Figure 5.13 shows the effect of increasing the SL time on the signal intensity. A plot of signal intensity as a function of the SL time will show a dependence with equation 5.1, this is displayed in figure 5.14. The calculated $T_1\rho$ decay constants from fitting equation 5.1 are displayed in table 5.1.

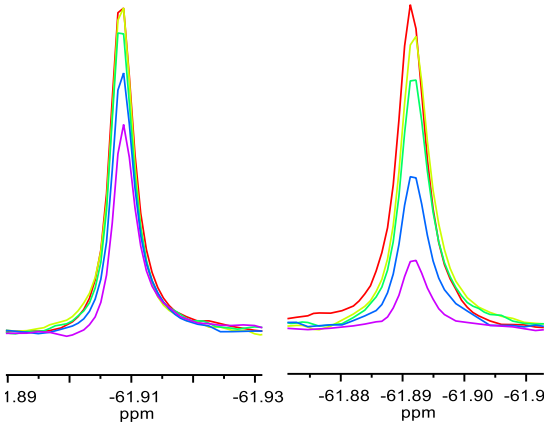
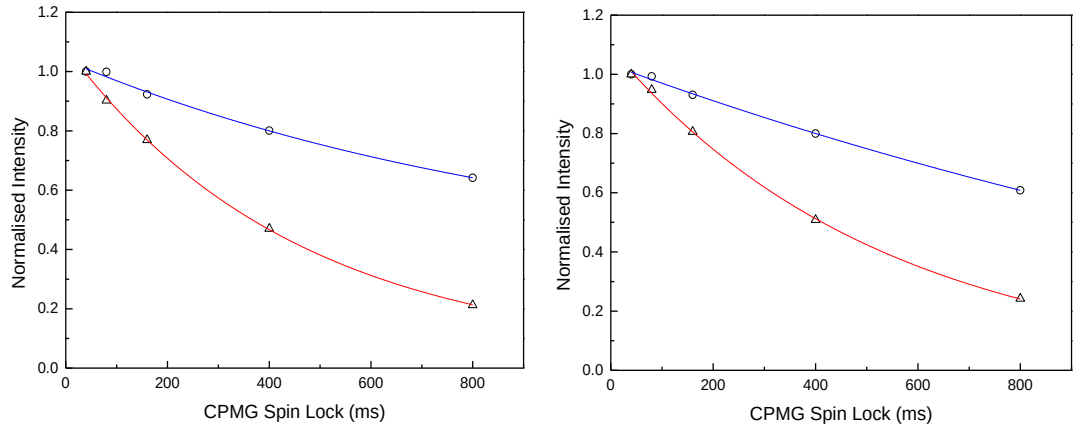


Figure 5.13: Effect of increasing the spin-lock time on the signal intensity. A decrease in intensity of the fluorine resonance is observable on increasing the SL time with the CPMG sequence. The left spectrum was acquired without protein, and the right spectrum was the result of an experiment containing 5 μM protein (CFL02 compound concentration was 100 μM). The corresponding SL times are: Red: 40 ms, Yellow: 80 ms, Green: 160 ms, Blue: 400 ms and Pink: 800 ms.

Compound	$T_1\rho$ (compound) (ms)	$T_1\rho$ (with protein) (ms)
CFL02	1680	493
CFL17	1507	532

Table 5.1: Calculated $T_1\rho$ decay constants for compounds (100 μM) in the presence and absence of 5 μM protein.

To demonstrate the use of $T_1\rho$ relaxation NMR as a potential ligand-observed screening method, the fragment pool containing the hits was also screened via this method. Figure 5.13 shows a superposition of the compound only and protein-containing pools acquired with a spin-lock filter incorporated. The length of the spin lock sequence was 400 ms and as can be seen from the spectra, there is a dramatic decrease in intensity of the two fragment hits (peaks 1 and 2). The fragments, peaks 3 and 4, show no difference in intensity of the two spectra indicating no-binding. Although this filter is a very sensitive method applicable to screening, incorporating a spin-lock into the experiment extended the total time of acquisition to 45 minutes per experiment. This is impractical for a high throughput, fragment screening approach if there are a large number of fragments present in the library.



CFL02 **CFL17**

Figure 5.14: Relationship of the spin lock time and the intensity of the fluorine resonance in the absence (blue curve) and presence (red curve) of 5 μM NusEΔ/NusB. Left plot: compound CFL02; right plot: compound CFL17. A single exponential was used to fit the intensity data and fitting was performed in Origin 9.0. It clearly shows a faster relaxation of the fluorine nucleus in the presence of protein indicating that the ligand is binding and that it adopts the properties of the protein.

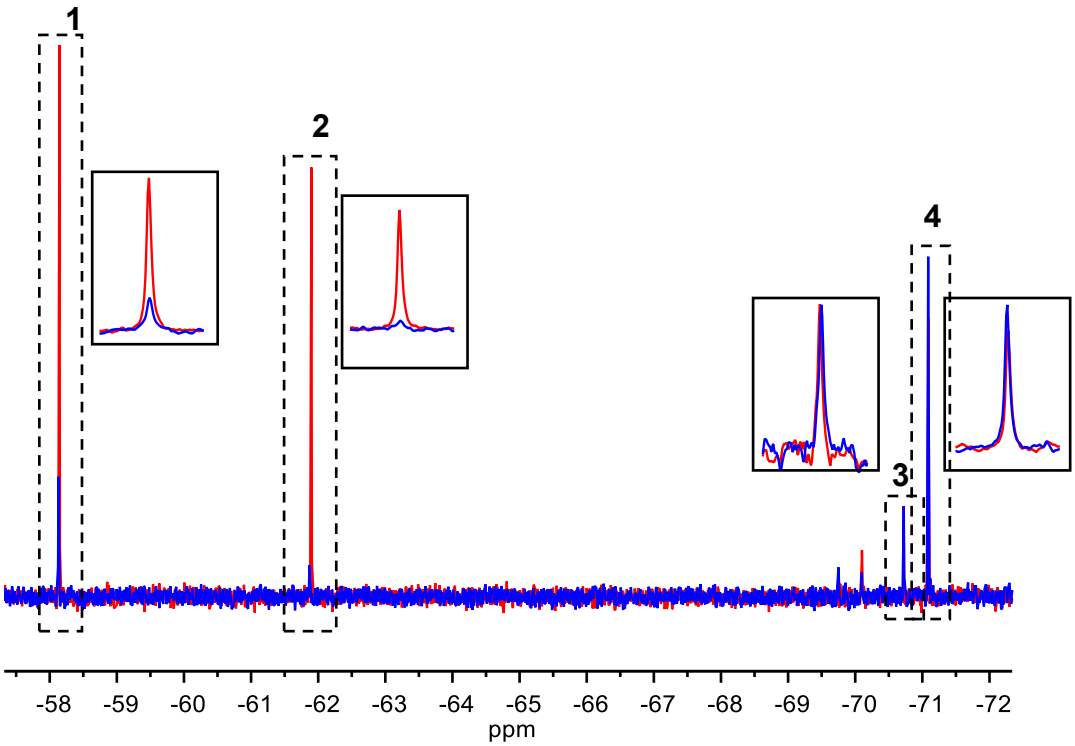


Figure 5.15: Superposition of two 1D ^{19}F NMR spectra with a spin-lock filter incorporated into the pulse sequence before acquisition; the length of the SL time is 400 ms. The red spectrum was acquired in the absence of protein, and the blue spectrum was acquired with 5 μM NusEΔ/NusB. Peak 1: CFL17. Peak 2: CFL02. Peak 3: CFL18. Peak 4: CFL19.

5.2.3. K_D determination by Differential Chemical Shift Perturbation (dCSP)

Determining the dissociation constant of a target binding hit compound is a critical parameter for selection of compounds for cellular testing, for structural biology and for medicinal chemistry. The binding of a ligand to a macromolecule is an equilibrium condition between free and bound states. The equilibrium ligand binding activity is therefore usually expressed as an affinity, K_D, of the ligand and protein to each other. NMR methods for the determination of dissociation constants are extensively covered in the literature. The use of ligand-observed NMR methods in drug discovery has led to much interest in binding affinity determination by NMR. The most traditional ligand-observed methods require that the ligand is in fast exchange (see Chapter 4) and in excess of the protein concentration, to satisfy the condition $[L]_0 >> [P]_0$ and, hence, $[L]_{free} = [L]_0$.[172] The difficulty here is that the observable entity in the experiment is the ligand, therefore usually the protein would be kept at a constant concentration. However, due to the relatively high concentrations needed for ligand excess, the solubility of fragments becomes an issue.

Jordan *et al* introduces a method for K_D determination by ^{19}F NMR which cannot be achieved with ^{1}H-based methods.[153] They call this 'Differential Chemical Shift Perturbation' (dCSP), which utilises the changes in the ^{19}F chemical shifts at two different concentrations of ligand. This implies that the change in chemical shift is proportional to the fraction of the bound ligand ($p_b = [PL]/[L]_0$). Provided that the total ligand concentration is in excess of protein, the K_D is given by equation 4.10 (see chapter 4.1.2):

$$K_D = \frac{\gamma[L]_1 - [L]_2}{1 - \gamma}$$

Equation 4.10

Where $[L]_1$ and $[L]_2$ are the concentrations of ligand and γ is the differential frequency shift (see chapter 4.1.2.1 for description). Samples were prepared with $[L]_1$ and $[L]_2$ of 100 and 200 μM respectively (for each compound) and protein concentration was kept constant at 5 μM (20:1 and 40:1 molar ratios). The experiment was performed in duplicate and for CFL02, the γ constant was determined to be 1.17 from the ratio of the frequencies, therefore K_D was estimated to be 463 ± 8 μM (this is shown in figure 5.16). For CFL17, the γ constant was determined to be 1.14, therefore the K_D was calculated to be 591 ± 2 μM. The K_Ds are approximately in the same range and can be considered good affinities and starting points for fragment hits.[213]

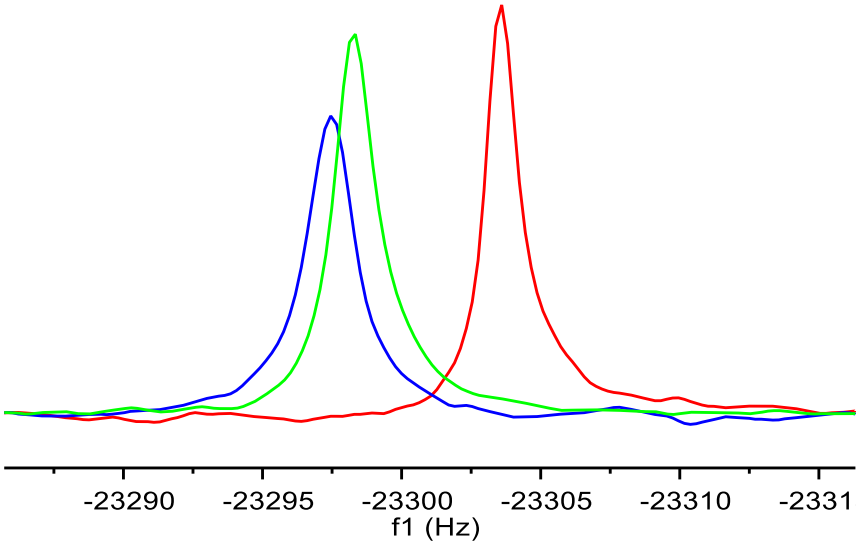


Figure 5.16: Change in Hz at two different ligand concentrations of CFL02. $[L]_1$ = 100 μM (blue curve); $[L]_2$ = 200 μM (green curve); γ = 1.17, therefore using equation 5.2.2, K_D = 463 ± 8 μM. The red curve is the reference peak of CFL02 that contains no protein.

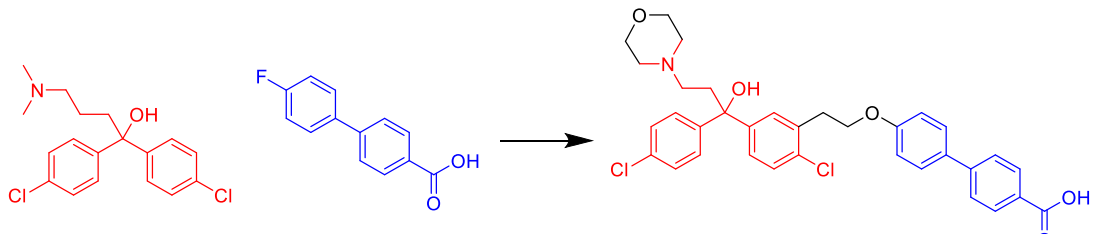
5.2.4. Fragment development

In the early drug discovery pipeline, the stage after hit discovery, confirmation and validation is the hit-to-lead phase. This is a period where hit fragments undergo optimisation into promising lead compounds.[214] Following hit validation and testing for artefacts, optimisation can be carried out by testing analogues (hit expansion) to improve and/or maintain affinity of the original compound. Readily available fragments can generally be sourced from commercial vendors enabling "SAR by catalogue".[164] One approach is to use sub-structure searches, where a fragment `core` is defined and different chemical moieties are displayed around it.

5.2.4.1. Fragment linking

Advancing fragments in the absence of a structure is a major challenge. The early work involved in the development of the fragment-derived drug Venetoclax, for example, relied only on NMR structures. Similarly, X-ray crystallography was unsuccessful against MCL-1[215], however, NMR-based models allowed efficient fragment advances. Also, *in silico* modelling continues to improve and has recently succeeded in linking fragments into a nanomolar inhibitor of Jumonji histone demethylases.[216] There are few examples of advancing fragments without structural information[217,218] however, it is worth noting that drug optimisation was

achieved without the benefit of structural information decades ago. Fragment linking is an appealing strategy that involves linking two, or more, low affinity binders to increase potency. This has been achieved several times, a notable example is fragment hits against Bcl-2, a regulator of apoptosis. The researchers identified 2 fragments with μM affinity from a protein-detected NMR screen of 17,000 compounds.[219] Structural NMR work showed the fragments bind simultaneously and linking them increased the affinity to the high nanomolar range (figure 5.17).



Compound 5.10
K_D = 20 μM

Compound 5.11
K_D = 300 μM

Compound 5.12
K_D = 0.22 μM

Figure 5.17: Compounds 5.10 and 5.11 were identified from an NMR screen with modest affinities to Bcl-2. Fragment linking led to compound 5.12 with high nanomolar affinity.[219]

Show in figure 5.17 is a good example of a fragment linking approach to increase both affinity and specificity towards a target protein. Therefore, it was hypothesised that if the fragments identified to NusE/NusB bind close enough in space, they may be amenable to a fragment linking approach. The Nuclear Overhauser Effect (NOE) is a physical phenomenon in which nuclear spin polarization is transferred between nuclei that are close in space, rather than through chemical bonds. This gives rise to cross correlation peaks in 2D NMR spectra and the rate of build-up of the NOE (i.e. the cross relaxation rate) is proportional to: $1/r^6$, where r = the internuclear distance.[220] It has recently been extended to observing protein-mediated interligand cross correlations.[221] This technique, known as Inter-Ligand Overhauser Effect (ILOE) relies on the fact that if two ligands interact simultaneously with the protein in adjacent binding sites, then ILOEs will be observed.[222]

The two compounds both contain a trifluoromethyl group, (CF_3) it was therefore considered to explore them in ^{19}F homonuclear NOESY experiments to decrease spectral complexity. Figure 5.18 illustrates a ^{19}F-^{19}F (^1H-decoupled) NOESY experiment for the two CF_3 containing fragments. However, no ^{19}F-^{19}F NOE cross-correlation was detected, indicating that either the compounds bind at different sites far apart (> 5 Å) or compete for the same binding site. This suggests that the compounds cannot be linked by this strategy. However, it was also

hypothesised that, no ILNOE (Inter Ligand Nuclear Overhauser Effect) was observed due to the distance between the CF_3 groups, therefore a second ^{1}NOESY experiment was acquired to observe the ^{1}H-^{1}H NOEs (figure 5.19).

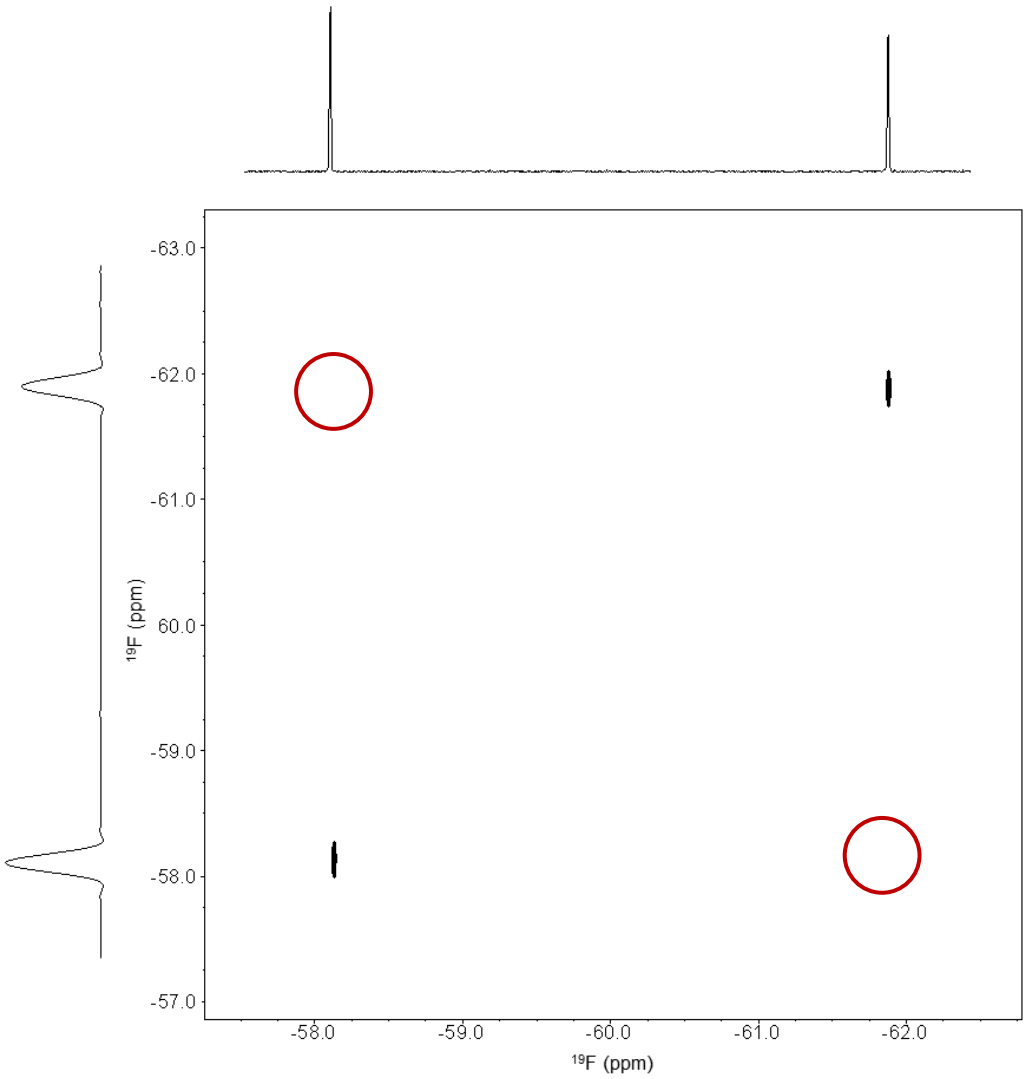


Figure 5.18: This figure shows a 2D homonuclear ^{19}F-^{19}F-NOESY spectrum of compounds CFL02 (-58 ppm) and CFL17 (-62 ppm) (compound concentration is 100 μM and protein concentration is 5 μM). ^{19}F-^{19}F ILNOESs are not observed between the two fragments. This would be indicated by a cross correlation peak in the indicated circles (red).

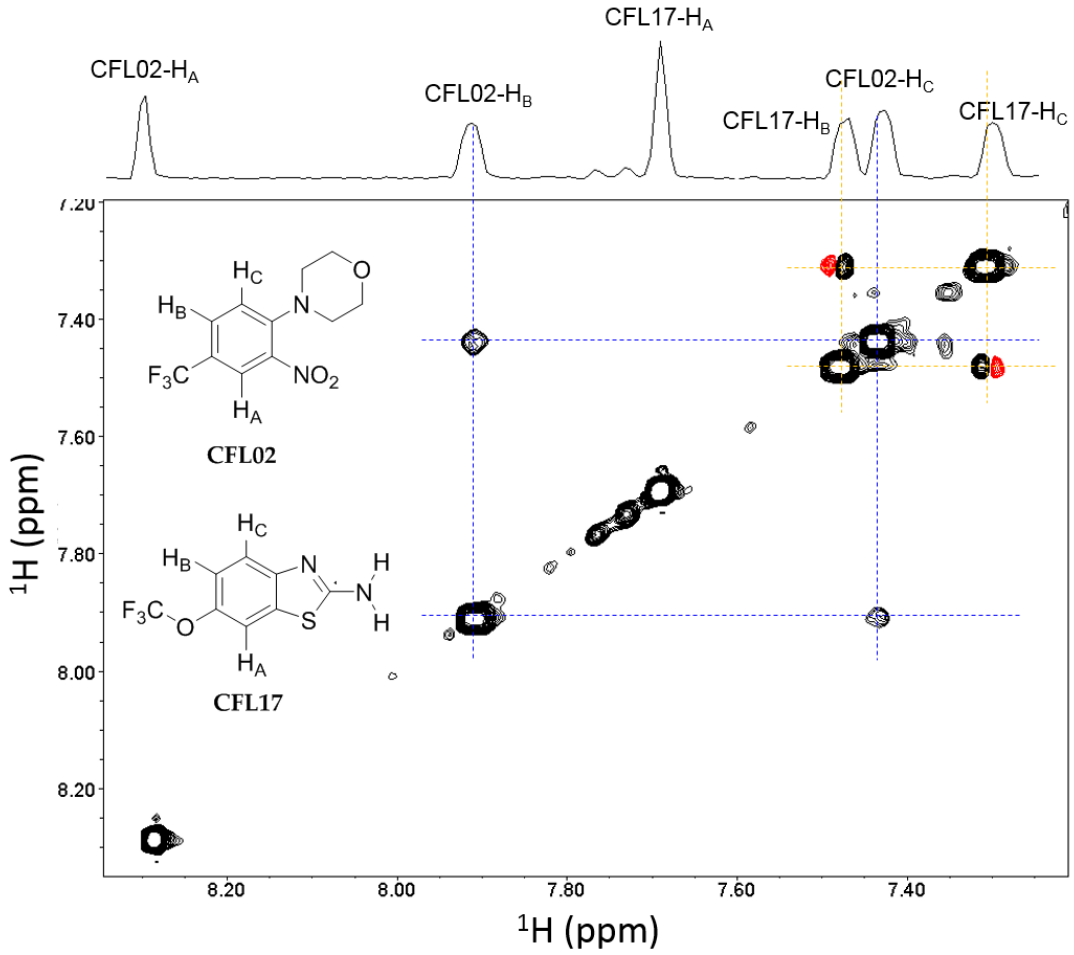


Figure 5.19: 2D ^{1}H-NOESY spectrum of a mix of two hit fragments (100 μM) in the presence of NusEΔ/NusB (10 μM). It shows the aromatic region of compounds and the cross correlation between aromatic protons. The blue square indicates an NOE between the two protons (B & C) of CFL02. The orange square indicates an NOE between the two protons (B & C) of CFL17.

As shown in Figure 5.19, there are no intermolecular NOE cross correlation peaks between protons of both fragments. Again, this could indicate one of two possibilities: 1) that these fragments do not bind to NusEΔ/NusB close together in space (> 5 Å apart), 2) they bind at the same site to NusEΔ/NusB, which means that they are competitive binders. However, the NOESY spectrum also shows a positive cross relaxation peak between protons of compound CFL02, this indicates a negative relaxation rate. This situation is encountered in the slow-tumbling or spin diffusion limit for large molecules, such as proteins. Therefore, it can be concluded that these fragments adopt the same, slow tumbling properties as the protein, which is another indication of ligand binding. [223]

In the following section the optimisation of the fragment hits by testing compound analogues to increase the binding affinity is described.

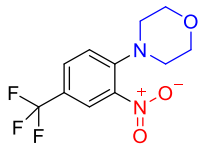
5.2.4.2. Fragment Optimisation by Bioisosteric Replacement

Bioisosteres are chemical functional groups that share similar physical and chemical properties which produce a broadly similar biological effect of the molecules containing them. The rationale behind exchanging bioisosteres on molecules is to enhance the biological effect without drastically changing the chemical structure. The replacement bioisostere can modify the pharmacokinetic properties of the lead compound, reduce toxicity and alter the bioavailability. An example of some simple, frequent replacements are shown in table 5.2.

Chemical Functional Group	Replacement
-H	-D, or -F
$-CH_3$	$-NH_2$, -OH, F, Cl
-Br	iPr, $-CF_3$
$-CH_2-$	-NH-, O, S
C	Si

Table 5.2: Table illustrating common bioisosteric replacements for chemical functional groups.

Compound CFL02 has two obvious chemical moieties for exploring changes in the biological activity, these are the nitro group and the morpholine group (figure 5.2.15). Nitro containing compounds are organic molecules that contain one or more nitro functionalities ($-NO_2$). Nitro groups are generally avoided by medicinal chemists owing to their relative reactivity and potentially toxic metabolites. They are readily reduced *in vivo* to hydroxylamines and nitrosamines which are highly reactive and known carcinogens.[224]



CFL02

Figure 5.20. Chemical structure of fragment CFL02 with the nitro-functional group (highlighted in red) and the morpholine group (highlighted in blue).

Although instances of nitro containing compounds are rare in nature, a well-known example of a naturally occurring nitro compound is chloramphenicol (Figure 5.21). Chloramphenicol is a broad spectrum antibiotic useful for treatment of a number of bacterial infections including: conjunctivitis, meningitis, plague, cholera and typhoid fever. It's mechanism of

action is by preventing the peptidyl transferase activity of the bacterial ribosome. It specifically binds A2451 and A2542 residues in 23S rRNA of the 50S ribosomal subunit, preventing peptide chain elongation.[225]

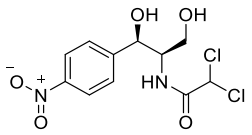


Figure 5.21: Chemical structure of the broad-spectrum antibiotic chloramphenicol.

Also, Zeng *et al* describe the development of a class of small molecules inhibiting the HIV-1 Tat and coactivator PCAF in which a nitro group on an aniline ring is essential for the binding activity.[226] An example of a nitro bioisostere replacement for drug discovery comes from Merck, who ran a HTS campaign against the ROM-K potassium ion channel.[227] They found a single hit compound that contained two aryl nitro groups. However, they followed this lead up and identified several bioisosteric replacements for the nitro group (shown in figure 5.22).

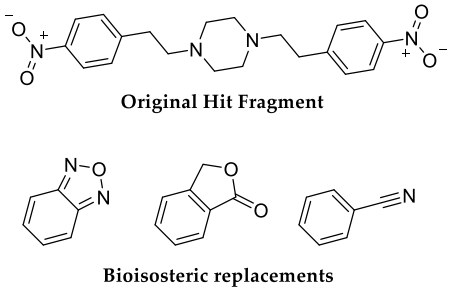


Figure 5.22: Chemical structures of the hit compound and several bioisosteric replacements as reported by Merck (2012). Figure adapted from [https://www.cambridgemedchem-consulting.com/resources/].

To optimise the fragments, a bioisostere replacement approach was used to generate similar compounds with enhanced affinity. The SwissBioisostere database contains information on specific molecular replacements and their effect in biochemical assays.[228] A query of the nitro group and morpholine groups in the database, and filtering for compounds with LogP < 0 (i.e. better aqueous solubility) returns a list of 254 and 377 possible replacements respectively. The top 4 from each query, ranked by frequency (i.e. number of times replacement occurs in database), are shown in table 5.3. This table also shows the percentage of biochemical assays in which the replacement was a "better" effector of the assay. For example, in 45% of assays where a nitro group was replaced with an amine or hydroxyl group, this resulted in a better 'output' from the assay. A structure search in eMolecules for commercially available

fragments containing these replacements was performed and it was found that replacements 1, 2 and 4 (of the nitro group) and number 4 (of the morpholine group) were available at a reasonable price. These compounds were purchased for testing against the NusEΔ/NusB protein complex (see sppendix 8.4.2).

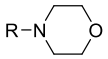
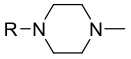
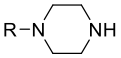
Query	No.	Replacement	Frequency	% of assays better
	1		1015	45
	2		792	27
	3		541	45
	4		298	33
	1		770	31
	2		409	40
	3		349	35
	4		234	38

Table 5.3: Result of a SwissBioisostere search query on the nitro and morpholine groups present in compound CFL02. The table is ranked by frequency.

5.2.4.3. Structure Activity Relationship Analysis by 2D Chemical Structure

A second approach to derive analogues for SAR screening was to use 2D structural similarity searches of the initial hit fragments. As described in chapter 3, similarity searching using the FCFP4 fingerprints was performed by Dr Steven Shave in the Auer group. Due to its highly decorated structure, few 3D similar compounds could be found for CFL02. Prioritised fragments are ranked on their USRCAT score (a measure of the 3D similarity) and shown in table 5.4 and 5.5. For CFL02, there were 4 compounds selected for purchasing, however, compounds 1 and 4 were prioritised for screening (table 5.4). These compounds tended to explore SAR around the morpholine moiety with complete removal of the nitro group. For CFL17, again 4 compounds were prioritised for purchasing, however, analogues of this compound was very expensive, therefore, only one fragment analogue was purchased for testing (compound 1, table 5.5). This compound contains a trifluoromethylthio group (-SCF_3) in place of the trifluoromethoxy group (-OCF_3).

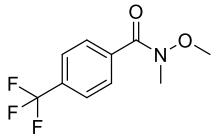
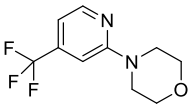
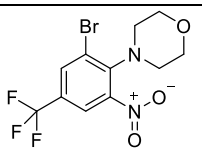
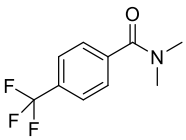
Number	2D Structure	Information
1		USRCAT Score:5.29
2		USRCAT Score: 6.60
3		USRCAT Score: 7.04
4		USRCAT Score: 7.13

Table 5.4: Ranking of fragment analogues to CFL02 prioritised for purchasing. The USRCAT score is a measure of 3D similarity, the lower the score, the more similar the compound is to the query. Compounds 1 and 4 were selected for purchasing.

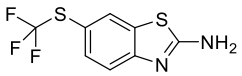
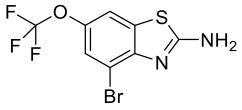
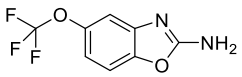
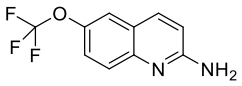
Number	2D Structure	Information
1		USRCAT Score: 4.92
2		USRCAT Score: 8.43
3		USRCAT Score: 9.40
4		USRCAT Score: 14.11

Table 5.5: Ranking of fragment analogues to CFL02 prioritised for purchasing. The USRCAT score is a measure of 3D similarity, the lower the score, the more similar the compound is to the query. Compound 1 was selected for purchasing.

The fragment analogues were screened by ^{19}F-NMR CSP as before, at 100 μM concentration in the absence and presence of 10 μM NusEΔ/NusB protein complex. The change in chemical shift of the fluorine resonances are given in figure 5.23. As the response of the fluorine

resonance is proportional to the fraction of bound ligand (p_b = [PL]/[L]$_0$) the largest change in fluorine resonance can indicate a tighter binding ligand. As can be seen from figure 5.23, analogues CFL02f and CFL17a (highlighted in green) both give larger resonance shifts than their corresponding initial hit compound. It can also be seen that changes to the morpholine moiety almost completely destroy the binding (CFL02c and CFL02d). This could indicate that this functional group is necessary for the binding effect. Replacement of the nitro moiety also has a detrimental effect on the binding interaction. Ranking the nitro bioisostere replacement by chemical shift results in: NO_2 > CN > NH_2 > COOH. Figure 5.24 shows the 1D ^{19}F NMR spectra for each analogue fragment in the presence of the NusEΔ/NusB complex and the change in ^{19}F chemical shift.

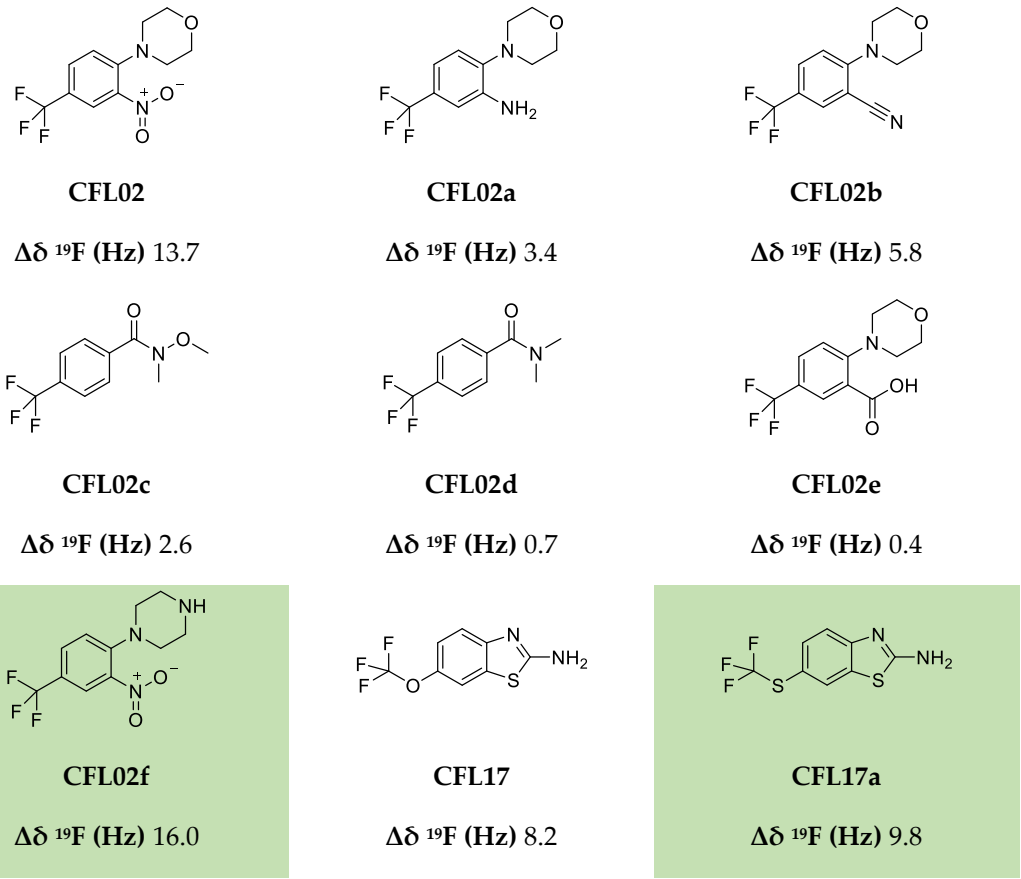


Figure 5.23: Chemical structure of each analogue purchased designed to increase/maintain affinity. Also shown is the change in chemical shift of the fluorine resonance (Δδ ^{19}F (Hz)) on addition of 10 μM NusEΔ/NusB. The highlighted fragments show a larger change in chemical shift on addition of protein than their corresponding initial hits (i.e. CFL02f > CFL02 and CFL17a > CFL17).

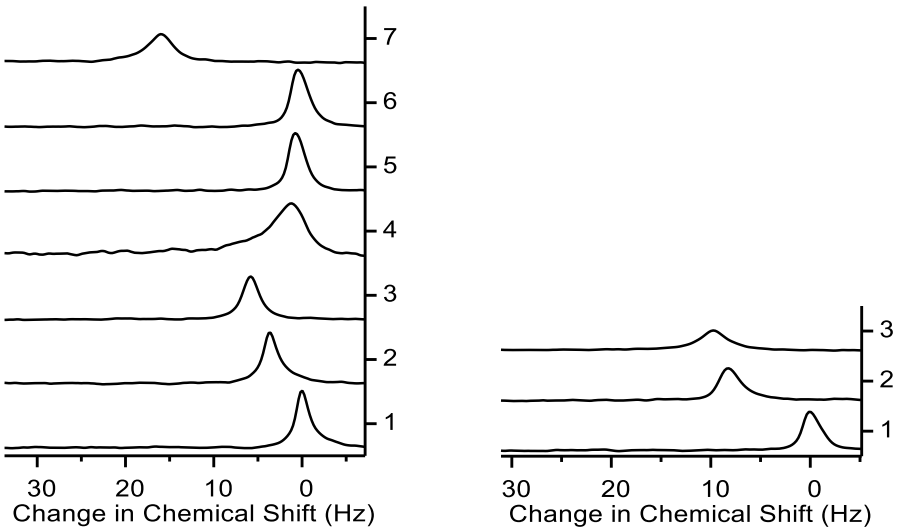


Figure 5.24: 1D ^{19}F NMR spectra of CFL analogue series. The spectra illustrate the change in ^{19}F chemical shift ($\Delta\delta$ Hz) for each analogue fragment on addition of 10 μM NusEΔ/NusB. Trace 1 is the reference peak for each compound in the absence of protein. Left: CFL02 series, 2 – 7: CFL02a – CFL02f. Right: CFL17 series, 2 – 3: CFL17 and CFL17a respectively.

The next step was to confirm the binding interaction and measure the affinities of the fragment analogues by application of a secondary label-free affinity technique.

5.2.5. Quantitative Equilibrium Microdialysis (qμD)

5.2.5.1. Theory

Equilibrium dialysis is well-established as a biochemical technique, and important for the characterisation of protein-ligand interactions. It is attractive due to its inherent physical simplicity and ability to work with non-labelled ligands and proteins (avoiding, e.g. fluorescence and radiolabelling). The general objective is to measure the affinity (dissociation constant, K_D) of small molecule ligands to macromolecules. Typically, this is exemplified indirectly as in a mixture of ligand and protein, the free ligand concentration is difficult to determine. Therefore, the theory behind equilibrium microdialysis is that, if the free ligand can be dialysed through a semipermeable membrane, then the free ligand concentration can be measured easily (illustrated in figure 5.25).

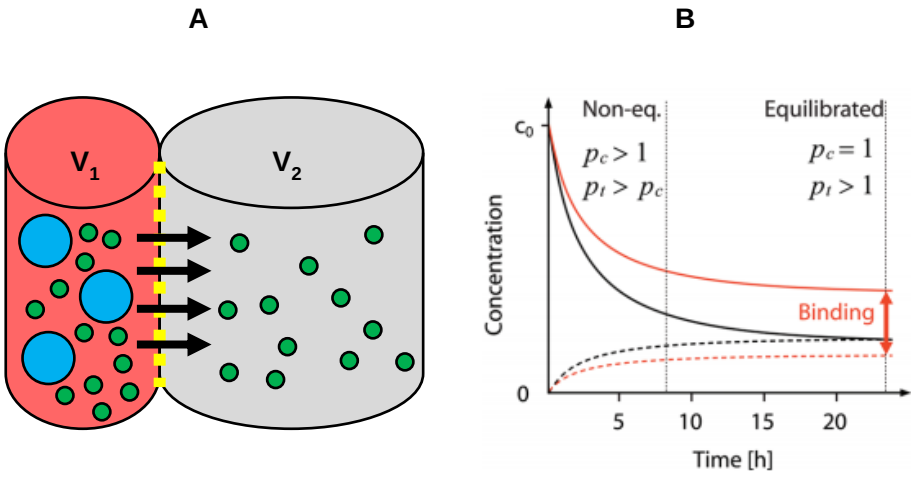


Figure 5.25: The theory behind the microdialysis experiment. **A)** A schematic of the microdialysis experiment. The red chamber volume (V_1) is 100 µl and the white chamber volume (V_2) is 300 µl. Blue circles represent protein molecules and green circles represent small molecule ligands. The yellow dashed line is a semi-permeable membrane with a MWCO of 8 kDa. **B)** A time course of the compound concentrations in the red chamber (solid lines) and white chamber (dashed lines). The black lines represent a control experiment without target protein, and the concentration is equal in each chamber after 24 h (p_c = 1). The orange lines illustrate the presence of target protein and after 24 h there is a difference in ligand concentration between the two chambers (p_t > 1). Figure B reproduced from Weidemann.[229]

As described by Weidemann *et al* the ratio of ligand in the two compartments, V_1 and V_2, can be used to detect protein binding in compartment V_1.[229] Under ideal conditions, in a system without protein (p_c subscript c for control), the concentrations in each compartment will be equal (p_c = 1, figure 5.25). The protein and ligand are added to compartment V_1 (p_t subscript t for target containing experiment) and if the ligand binds to the target protein a certain fraction of compound will be retained in the protein compartment (p_t > 1). The free ligand concentration can then be measured in compartment V_2 with high accuracy. Weidemann relates the K_D to the partition coefficient (or ratio of ligand concentration in each compartment) via the following equation:

$$K_D = \frac{c_p}{p_t - 1}$$

Equation 5.1

Where c_p is the total protein concentration in the red chamber (V_1) and p_t is the partition coefficient in the experiment with protein (i.e. ratio of ligand between the chambers, r_o). However, this relationship is independent of the volume of each compartment, and it does not allow for the case if the compound does not fully equilibrate across the membrane even in

the absence of protein (i.e. $p_c \neq 1$). In the Auer lab, Dr Steven Shave, has established a new binding equation to correct for the situation when $p_c \neq 1$ (shown in appendix 8.4.3). This establishes a new equilibrium for when the compound does not fully equilibrate.

5.2.5.2. Microdialysis Results

The objective here was to measure the dissociation constants of the initial hit fragments and the analogues with improved affinity. Using an orthogonal, biophysical assay to validate the NMR dCSP data, quantitative microdialysis (abbreviated qμD) was chosen as an effective method. The method utilised the high accuracy LCMS-Orbitrap mass analyser to detect and quantify the fragment compounds. A simulation of the experimental conditions is shown in Figure 5.26. From this figure it can be seen that as the K_D of the interaction increases the concentration of ligand and protein in each chamber converges. Therefore low affinity binders are difficult to determine without high accuracy methods. Here it is important to note that the K_D is essentially calculated by measuring the free concentration of ligand, as opposed to measuring saturation curves.

A calibration curve for each compound was produced to calculate the limit of detection and the limit of linearity of the analyser (Figure 5.27) (see chapter 7.4.7 for materials and methods). This would aid in setting up the experiment to determine the optimum concentration of ligand and protein to gain the largest difference between the chambers. Compounds CFL02, CFL02f and CFL17 all produced excellent standard curves for concentration determination (appendix 8.4.3). However, compound CFL17a did not produce a reproducible signal in the mass analyser, and was therefore not amenable to detection by this method.

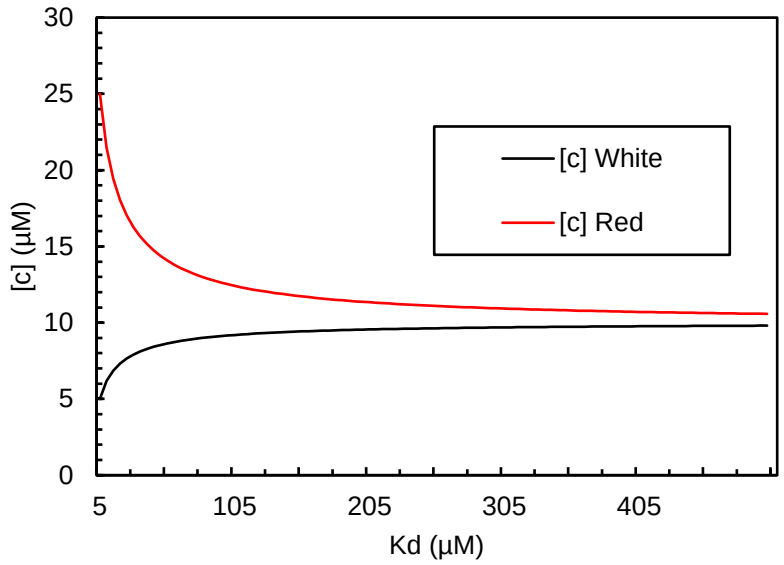


Figure 5.26: Simulation of the concentration of ligand present in each chamber for a given K_D. The protein concentration in the red chamber (V_1) is 40 µM. The ligand concentration over the whole system without protein is 10 µM ($p_c = 1$). [c] red = concentration of ligand in red chamber (V_1) and [c] white = concentration of ligand in white chamber (V_2).

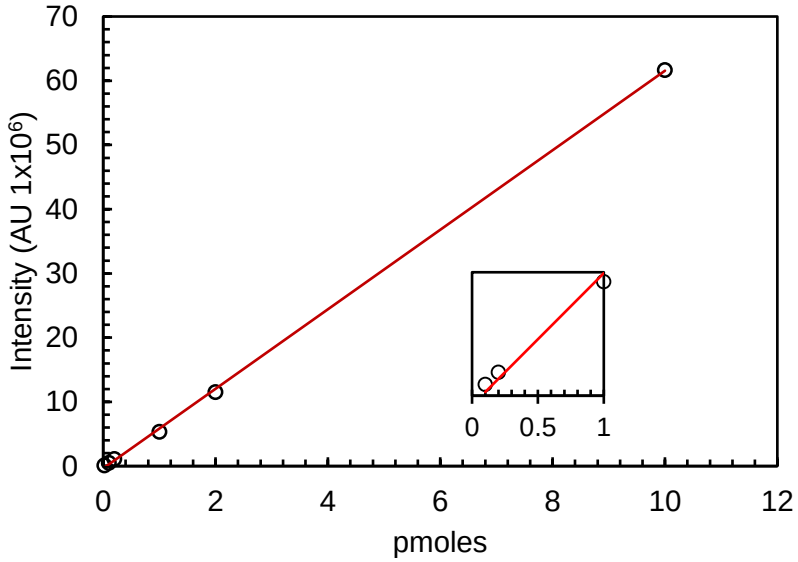


Figure 5.27: An example calibration curve for compound CFL02f detected on the LTQ orbitrap mass analyser. The limit of linearity is between 0.5 pmoles and 10 pmoles. This

The next step is to determine that compounds equilibrate sufficiently across the membrane. Initially, the hit fragments (CFL02 and CFL17) were tested for equilibration, however, these compounds did not give reproducible data and the calculated partition coefficient was concerning (table 5.6). The deviation of > 10% was considered unsuitable for accurate K_D

determination even with the adjusted K_D method where $p_c \neq 1$. This could be attributed to the compounds "sticking" to the membrane or the plastic walls of the sample chamber, this would result in a lower concentration than expected. Also, pipetting is of great importance here as the technique is very sensitive to changes in concentration. Compound CFL02f was the only compound that dialysed adequately and with reproducible results, that is, the standard deviation of the other fragments tested was too large to be reliable.

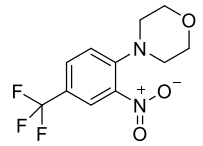
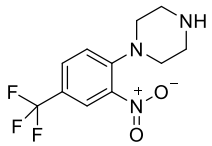
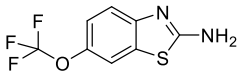
Compound ID	Structure	Mean peak intensity V_1 (10^6)	Mean peak intensity V_2 (10^6)	Partition coefficient
CFL02		2.39	2.06	1.16 ± 0.33
CFL02f		6.06	6.15	0.99 ± 0.03
CFL17		3.52	4.26	0.83

Table 5.6: The mean peak intensity detected for each compound in the red chamber (V_1) and white chamber (V_2) in control (protein free) microdialysis. The calculated partition coefficient (p_c) of each compound is also given.

The partition coefficient (p_t) for compound CFL02f was then assayed in the presence of NusEΔ/NusB and was found to be considerably higher than the control experiment (p_c), indicating protein binding (table 5.7). The p_t values, protein concentration and ligand concentration were then used to determine the K_D using both the Weidemann formula (equation 5.1) and the corrected S. Shave formula (appendix 8.4.3). Here, the K_D value given for this compound is comparable using both methods.

Compound	Mean peak intensity (V_1) (10^6)	Mean peak intensity (V_2) (10^6)	Partition coefficient	K_d (μM) (Weidemann)	K_d (μM) (Auer)
CFL02f	7.26	5.98	P_t = 1.21 ± 0.04	190 ± 35	171

Table 5.7: Results of the microdialysis equilibrium experiment with compounds CFL02f in the presence of NusEΔ/NusB. The peak intensities in the red chamber (V_1) and white chamber (V_2) following dialysis with protein are given as well the partition coefficient (p_t). The resultant K_D value using both formulas is also shown.

This assay proved uncomplicated for confirming the binding of one hit analogue identified in the ^{19}F NMR screen. However, the disadvantage with assaying small molecule compounds is establishing a sufficient equilibrium across the dialysis membrane, this proved problematic for 3 out of 4 compounds. Also, another limitation is the duration required for the system to reach equilibrium so it would be unsuitable for potentially unstable proteins. However, it validates the use of this method as a secondary binding assay for confirmation of fragment hits.

5.2.6. Bacterial Growth Inhibition Assay

The increasing emergence of new antibiotic resistant bacteria in hospitals and the community has led to rapid growth in strategic antibacterial drug campaigns. The rate of antibacterial resistance has greatly superseded the rate of discovery of new, effective drugs hence new lead compounds are desperately needed to respond to this. As our target is a bacterial PPI, it was relevant to assess the activity of the hit fragments identified in the initial ^{19}F NMR-based screen and the SAR analogues on bacterial *E. coli* cells.

A plate-based growth inhibition assay was devised to observe the effect of different concentrations of compound on bacterial growth. Here, the growth of *E. coli* was monitored by measuring the OD_{600} (optical density at 600 nm) at regular intervals over approximately 16 hours. Using 96-microwell plates and a compatible plate reader allowed for easy set-up and for testing several compounds and concentrations at once (see chapter 7.4.6 for materials and methods). The bacterial growth strain chosen was *E. coli* BL21(DE3), as this was readily available and has predictable growth properties. An initial test of growth conditions was performed to determine the initial optimum cell density. In addition, fragment stocks were stored in 100% DMSO therefore it was necessary to determine the effect of DMSO concentration on cell growth. It was verified that DMSO concentration has a negligible effect on the growth rate of *E. coli* up to a final concentration of 4% (figure 5.28). However, decreasing the initial concentration of cells (OD_{600}) increased the lag time of the growth curve as expected (see appendix 8.4.4). This was not an issue over ~16 hours, and the final parameters chosen were a starting OD_{600} of 0.001 and a final DMSO concentration of 1% (see chapter 7.4.6 for full materials and methods).

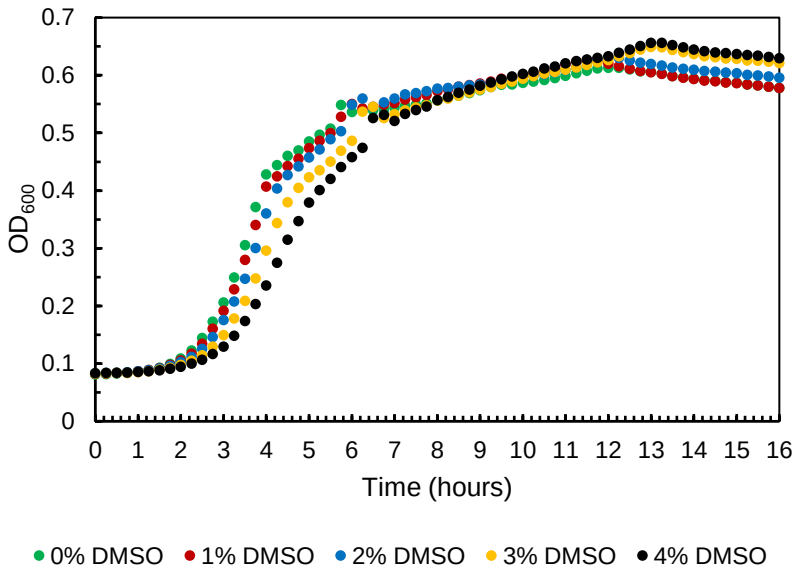


Figure 5.28: Effect of DMSO concentration on cell growth in the plate-based assay. The starting OD was 0.001 and the OD at 600 nm was measured every 15 minutes over a 16 hour period. There is a slight decrease in the rate of growth on the addition of increasing amounts of DMSO, It is shown that the effect of increasing DMSO concentration is not especially detrimental to the rate of cell growth.

A typical bacterial growth curve shows five distinct phases: lag phase, exponential growth, stationary phase, cell death and, finally, long-term stationary phase. Lag phase remains the most poorly understood growth phase, however, it is hypothesised that lag allows the adaptation of bacterial cells to exploit new environmental conditions.[230] Examining the growth curves from the hit fragments and SAR analogues, two different effects are observed. The benzothiazole containing CFL17 fragment series (CFL17 and CFL17a) causes an elongation of the lag phase for several hours, however, the growth rate is eventually restored to the same as the control (figure 5.29). This could be described as a partial bacteriostatic effect. An increase in lag phase of several hours indicates a requirement of the cells to adapt to a new environment. Importantly, the lag time increase shows a concentration dependence, i.e. lag time increases with increasing compound concentration. This strongly indicates a compound effect, negatively affecting bacterial growth and forcing bacteria to adapt to a new environment. However, there is no direct indication that this effect is due to inhibition of the NusE/NusB interaction pathway. Such a small fragment would most likely interact with several proteins involved in cell growth. However, as the NusE/NusB factors are essential for cell growth[206], knockdown of this pathway is not possible to test this hypothesis. An additional control experiment would be to carry out this assay on a eukaryotic cell line, such as yeast.

The NusB protein is specifically a prokaryotic system involved in transcription, and NusE shares little homology with interfamily ribosomal proteins.[210] Therefore, complete inhibition of this pathway would not be expected to incur an effect on a eukaryotic species.

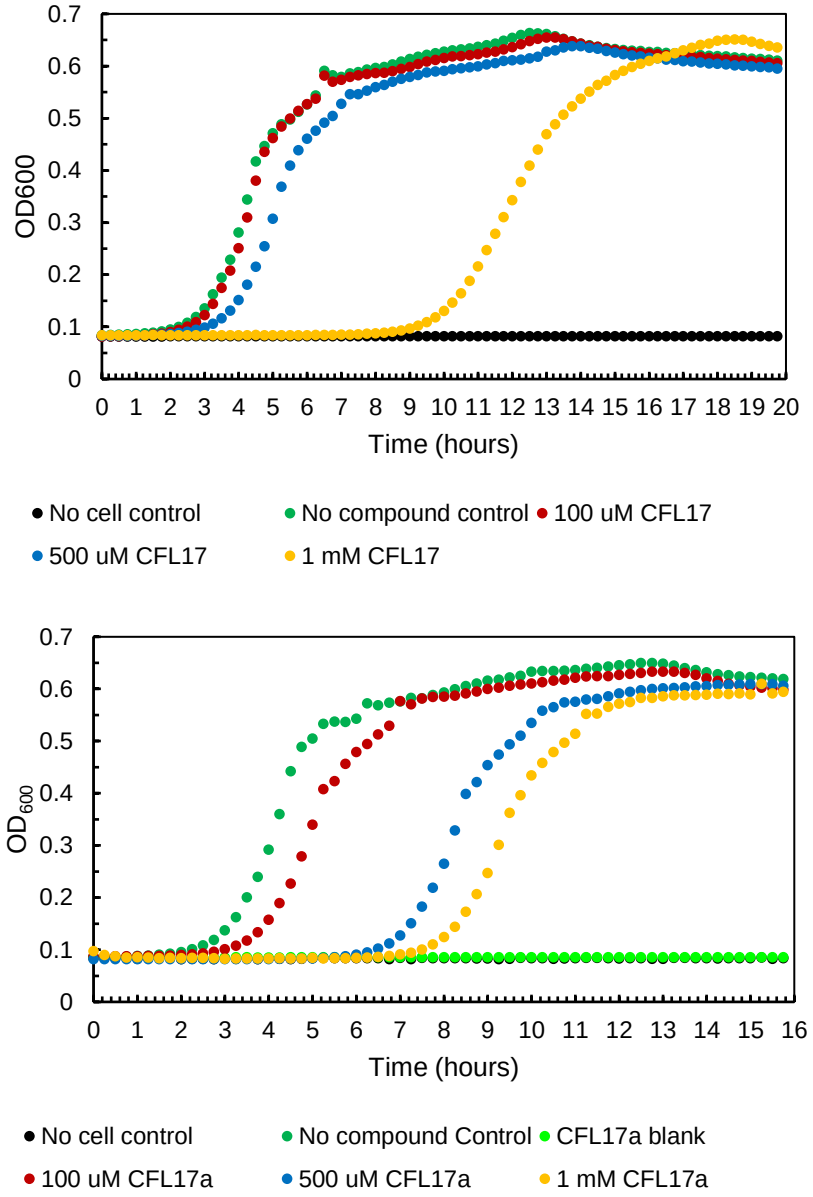


Figure 5.29: Effect of varying compound concentrations of CFL17 series on bacterial growth. Plot of the OD_{600} vs time (hours) in Microsoft Excel. Top: Initial fragment hit CFL17. Bottom: CFL17a (compound concentrations are indicated in figure legends). It can clearly be seen that on increasing the compound concentration an increase in the lag time is observed.

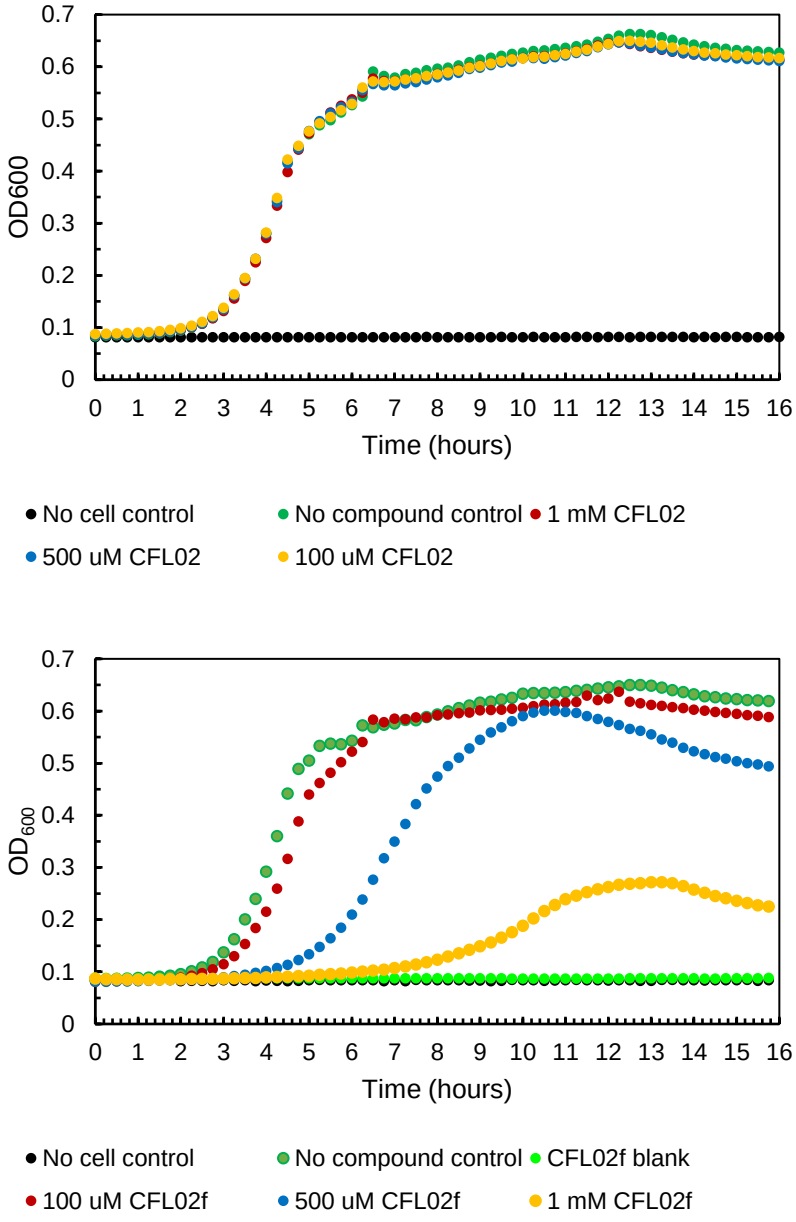
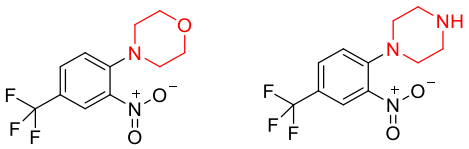


Figure 5.30: Effect of varying compound concentrations of CFL02 series on bacterial growth. Plot of the OD_{600} vs time (hours) in Microsoft Excel. Top: Initial fragment hit CFL02. Bottom: CFL02f (compound concentrations are indicated in figure legends). CFL02 shows little effect on bacterial cell growth.

Interestingly, the initial hit compound CFL02 did not show any activity on *E. coli* cells (figure 5.30, top) whereas the analogue CFL02f (figure 5.30, bottom) caused a dramatic decrease in cell growth at high (1 mM) concentration. This indicates the compound interacting with bacterial growth pathways. There is also a concentration dependence of each compound (CFL02f and CFL17a) as the effect is more pronounces at higher concentrations.



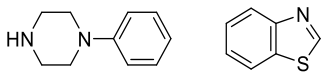
CFL02 **CFL02f**

Figure 5.31: Chemical structures of fragment hit CFL02 and analogue CFL02f showing the morpholine and piperazine structures highlighted in red.

The only difference between the two fragments is replacement of the morpholine moiety (in CFL02) with a piperazine ring (in CFL02f) (figure 5.31). A morpholine group is commonly used in medicinal chemistry to increase the aqueous solubility and, hence, oral bioavailability of lead drug candidates.[231] This could explain how, due to its lower lipophilicity, it is unable to cross the *E. coli* cell membrane and cause an effect.

5.2.6.1. Identifying the active fragment core

Interestingly, the fragment hits and SAR compounds to the NusEΔ/NusB complex contain recognised core structures that are present in many active pharmaceuticals. CFL02f contains a substituted phenylpiperazine structure (Figure 5.32; **1**) which is a known moiety in several drug derivatives. These drugs have a number of targets for a range of pharmacologically relevant illnesses, including bacterial infections, shown in table 5.8. The hit compound CFL17, and the thioether analogue (CFL17a), both contain the central benzothiazole core (Figure 5.32; **2**). This particular core is a useful moiety in several drug compounds targeting the central nervous system,[232] however it also finds use in antibacterial discovery.[233,234] This fragment compound (CFL17), as selected via the clustering design method, is actually a known drug compound, riluzole (marketed as Rilutek and Teglutik). It is currently used to treat amyotrophic lateral sclerosis (ALS). It's mechanism of action occurs through inhibition of TTX-sensitive sodium channels, which are associated with damaged neurons.[235]



1 **2**

Figure 5.32. Chemical structures of the fragment cores possibly responsible for the antibiotic activity. 1) phenylpiperazine, 2) benzothiazole.

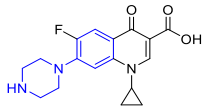
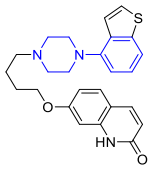
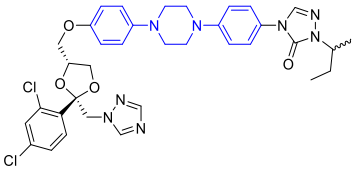
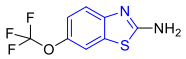
Drug	Structure	Pharmacology
Ciprofloxacin		Broad spectrum antibiotic of the fluoroquinolone class. It is Active against both Gram - positive and Gram-negative bacteria. The molecular target for this small molecule is DNA gyrase and topo-isomerase IV.
Brexpiprazole		Used in the treatment of schizophrenia and as an adjunctive for depression. This small molecule is a partial agonist for the dopamine D_2 receptor and is also a serotonin-dopamine activity modulator (SDAM).
Itraconazole		A broad spectrum antifungal agent, particularly against *Aspergillus*. It inhibits the fungal-mediated synthesis of ergosterol, via inhibition of lanosterol 14α-demethylase.
Riluzole		Used to treat amyotrophic lateral sclerosis.

Table 5.8: A range of drug compounds containing the similar moieties found in the fragment hits. Table is adapted from Trevor *et al*.[236]

Despite the low affinity of the hit fragment (590 μM, measured using NMR dCSP) against the NusEΔ/NusB PPI, CFL17 (riluzole) could, in theory, be modified and optimised to increase the potency against this difficult target. However, from a medicinal chemistry viewpoint it would be necessary to monitor side effects on the CNS. This is known as drug repurposing (also as drug repositioning) and purports the use of known drugs and compounds to treat alternative diseases. Two of the most famous examples of drug repurposing include: Pfizer's Viagra, initially studied for use in hypertension and angina pectoris, is now used for erectile dysfunction and pulmonary arterial hypertension, and Celgene's thalidomide, primarily described to alleviate morning sickness in pregnant women, it is now used as a first line treatment in multiple myeloma.

5.2.7. 2D HSQC NMR Experiments

The 2D HSQC (Heteronuclear Single-Quantum Coherence) experiment allows one to obtain a 2D heteronuclear chemical shift correlation map between directly bonded ^{1}H and any X heteronuclei (notably ^{13}C and ^{15}N with respect to protein NMR). This is the most standard experiment for protein NMR and shows a fingerprint of the backbone amide correlations. It is widely used as it involves irradiation and detection of ^{1}H spins, therefore offering high sensitivity and efficiency. It requires isotopic labelling of the protein with nitrogen-15 and involves the transfer of magnetisation from the backbone amide proton to the attached ^{15}N via J-coupling. The chemical shift is then evolved on the nitrogen before magnetisation is transferred back to the hydrogen for detection. The HSQC experiments were performed in collaboration with the Department of Biopolymers at the University of Bayreuth at their world leading NMR facility on a Bruker Aeon 1 GHz NMR spectrometer. They kindly supplied the ^{15}N labelled NusEΔ/NusB complex for K_D determination and chemical shift mapping. In this complex it is only the NusEΔ protein that is ^{15}N-labelled.

5.2.7.1. K_D Determination

The determination of binding affinity after hit generation is crucial for ranking compounds for selection and optimisation. For equilibrium ligand binding kinetics with respect to NMR spectroscopy, please see chapter 4.1.2.Using a ^{15}N-labelled protein, ligand binding is detected by perturbations to the chemical shift of backbone amide resonances. Figure 5.33 illustrates the titration of compound CFL02f into a solution of ^{15}N labelled NusEΔ/NusB. The CSPs of backbone amide resonances can clearly be seen for specific amino acids on increasing concentration of ligand. The amide resonances of NusEΔ have been previously assigned by the Biopolymer group.[237] The six active fragments (CFL02, CFL02a, CFL02b, CFL02f, CFL17 and CFL17a) in the ^{19}F NMR screen were used in titrations against the ^{15}N labelled NusEΔ/NusB protein for accurate K_D determination (see Chapter 7 for full experimental details and appendix 8.4.5 for HSQC spectra). Monitoring of the CSPs in the protein backbone allows one to create multiple saturation curves for each compound. A control experiment in which DMSO was titrated into the protein was also performed. This allowed for correction of CSPs that were affected by DMSO concentration (this is shown in figure 5.34). Also, multiple spectra of protein in the presence of 10% DMSO were collected over 72 hours to determine the stability of the protein in the presence of DMSO. After 72 hours, no protein degradation was visible by analysis of resonances (see appendix 8.4.5).

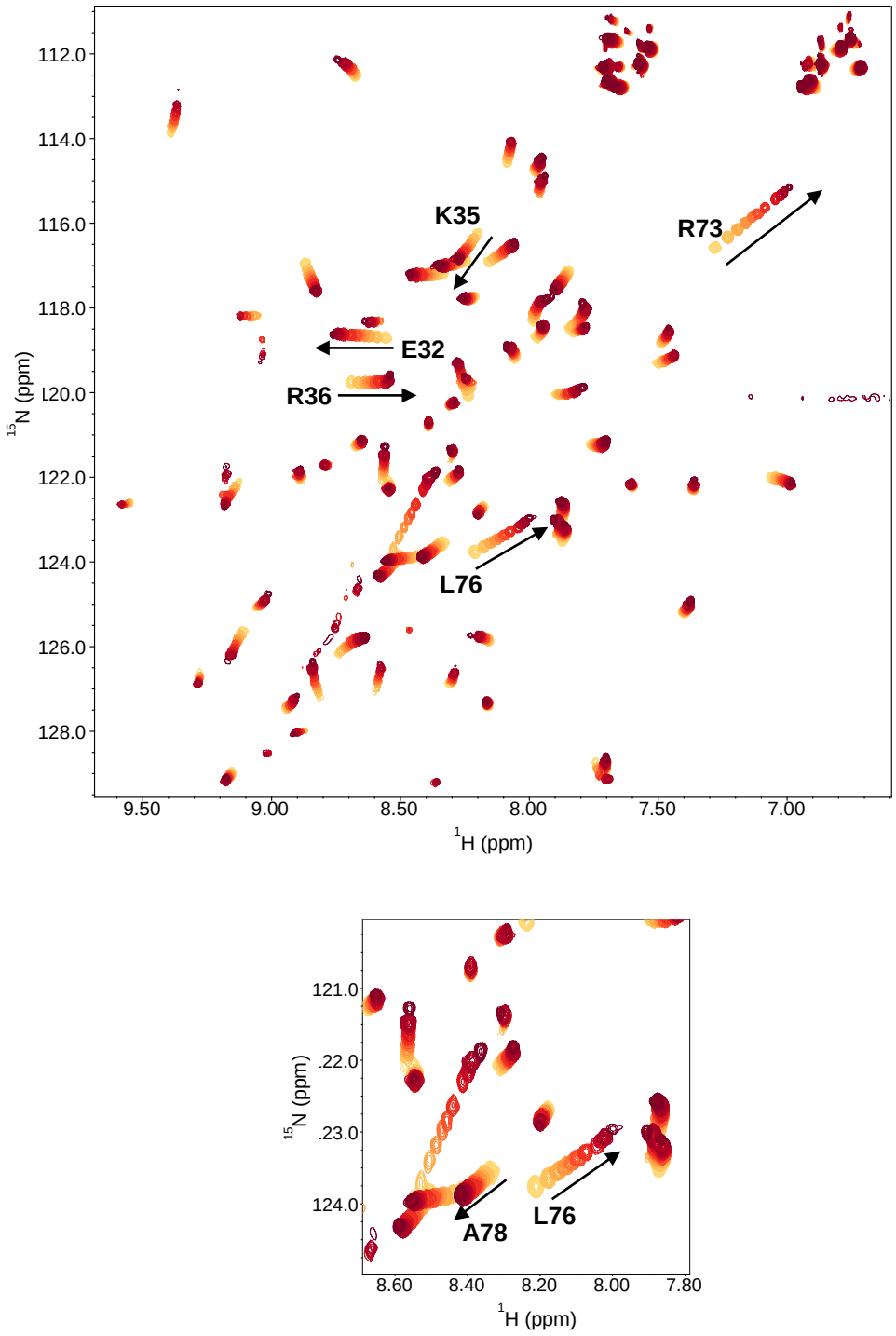


Figure 5.33: 2D ^{15}N HSQC spectrum of a series of titration points with compound CFL02f and ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex. From light to dark shows the change in chemical shift on increasing concentration of ligand. Top: Full spectrum of amide region. Bottom: Inset of residues L76 and A78. Concentration of ligand ($[L]_0$) ranges from 0 μM to 1.9 mM. 11 titration points were collected in total and the CSPs monitored.

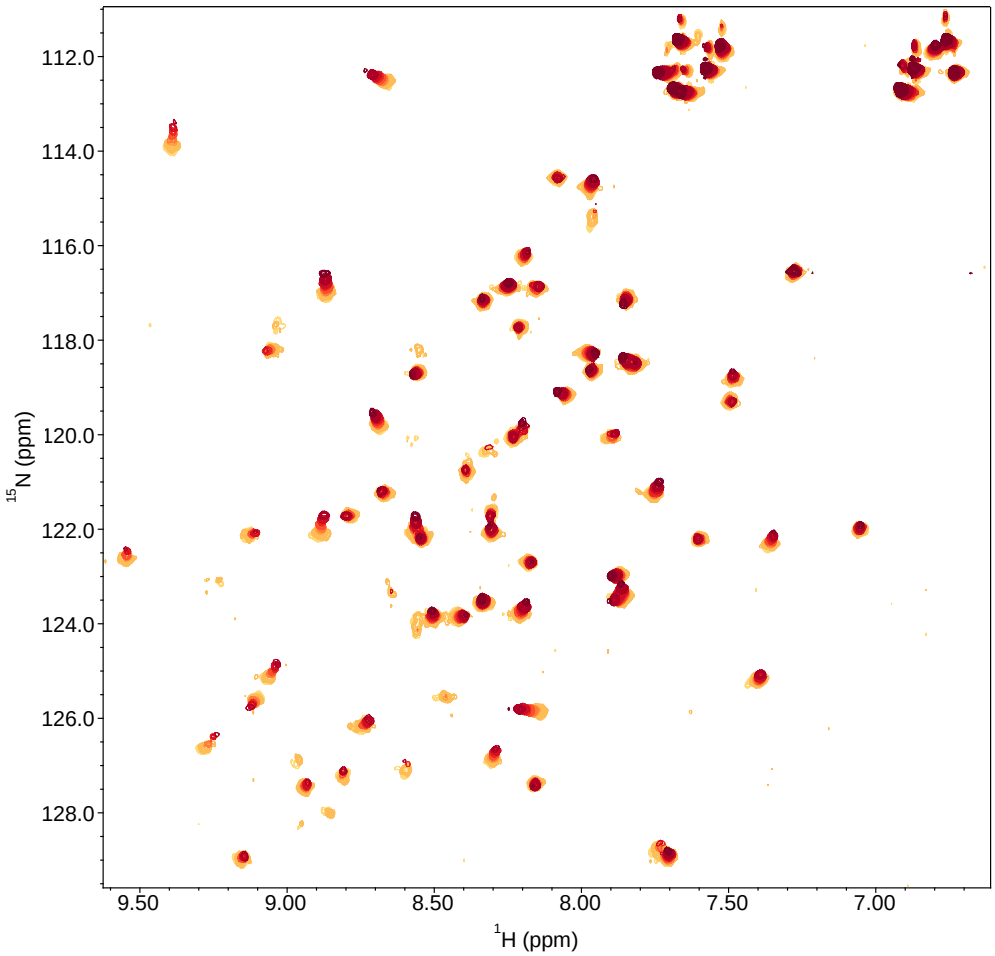


Figure 5.34: 2D ^{15}N HSQC spectrum of NusEΔ/NusB with DMSO titrations. From light to dark is increasing concentration of DMSO to a maximum 10% (equivalent to the final titration point with compound).

Figure 5.35 shows a titration curve for CFL02f. The normalised chemical shifts were calculated using equation 4.12 and the binding curve fitting analysis is described in Chapter 4. In theory, this analysis can be applied to each residue that changes over the course of the titration, therefore multiple curves can be acquired for each compound and the K_D values averaged. This was done for compound CFL02f, the standard deviation was deemed suitable that the K_D value from one residue gave an accurate representation of the dissociation constant. The calculated dissociation constants for multiple residues in the titration experiment with CFL02f are given in table 5.9. The fitting parameters and calculation of normalised chemical shifts are given in Chapter 7 (Materials and Methods).

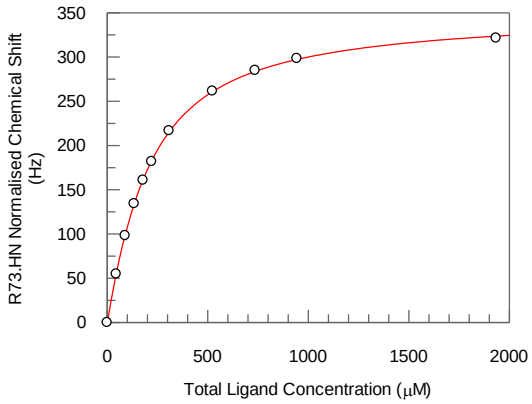
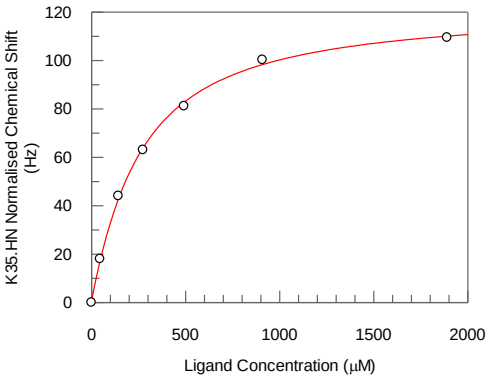


Figure 5.35: A plot of the saturation curve for compound CFL02f. The normalised chemical shifts were calculated for the R73 amide residue according to equation 4.11. The curve was fit to equation 4.12 using GraFit 7.0. The K_D from this curve is calculated to be 154.4 ± 2.7.

Residue	**K_D (μM)**
R73.HN	154 ± 3
E32.HN	167 ± 4
K35.HN	159 ± 4
R36.HN	124 ± 10
L76.HN	178 ± 3
Average	**156**

Table 5.9: Compound CFL02f K_D values for different residues that shift in the ^{15}N-HSQC spectrum. The .HN refers to the fact this is an amide nitrogen that is observed, i.e. R73.HN is the amide nitrogen of the arginine 73 residue.

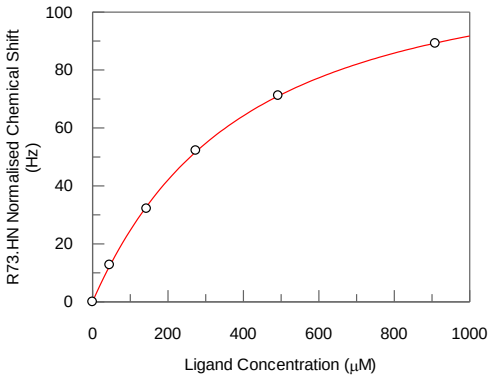
The calculated dissociation constants and titration curves for each compound are shown in figure 5.36. The curves are fit to equation 4.12 and the K_D is given in μM.



CFL02

Δ_{max} (Hz) 122.6 ± 2.2

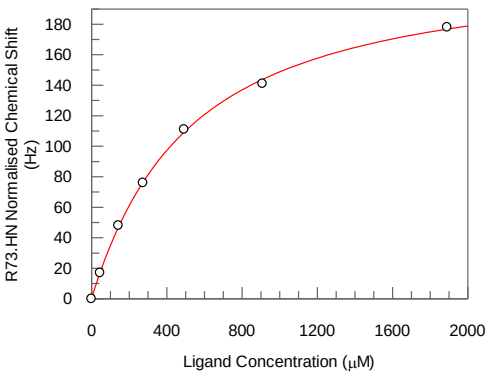
K_D (μM) 206.6 ± 16.1



CFL02a

Δ_{max} (Hz) 123.8 ± 2.0

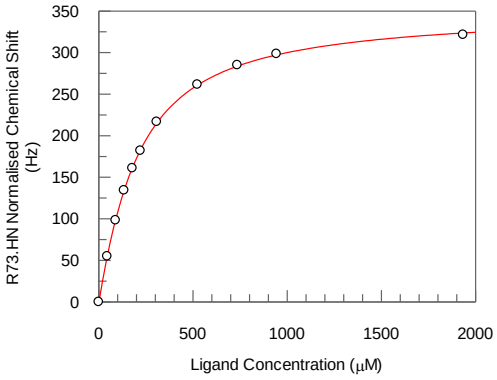
K_D (μM) 455.3 ± 25.9



CFL02b

Δ_{max} (Hz) 221.2 ± 4.0

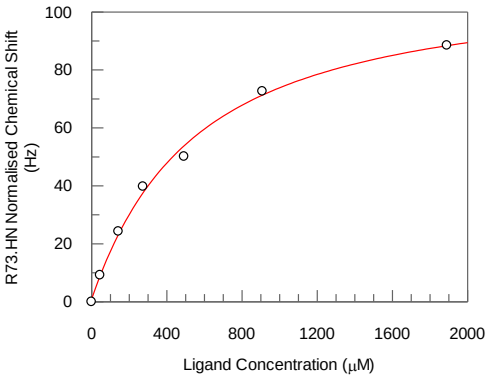
K_D (μM) 323.9 ± 14.9



CFL02f

Δ_{max} (Hz) 350.8 ± 1.4

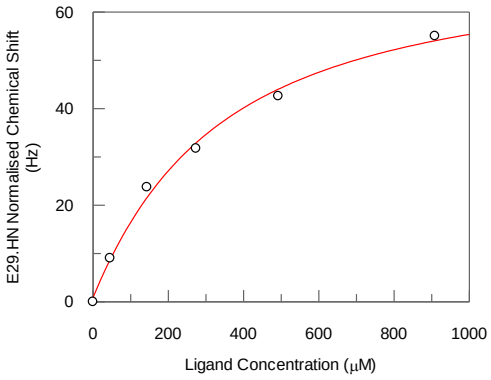
K_D (μM) 154.4 ± 2.7



CFL17

Δ_{max} (Hz) 111.9 ± 5.4

K_D (μM) 487.8 ± 72.0



CFL17a

Δ_{max} (Hz) 71.8 ± 5.5

K_D (μM) 277.7 ± 65.0

Figure 5.36: Titration curves for all compounds screened against ^{15}N NusEΔ/NusB. The Δ_{max} and K_D are used as a fitting parameters. Curves are fit using GraFit 7.0.

Compound	K_d (μM)
CFL02	206 ± 16
CFL02a	455 ± 26
CFL02b	323 ± 15
CFL02f	156 ± 3
CFL17	487 ± 72
CFL17a	277 ± 65

Table 5.10: Summary of calculated K_D values by 2D ^{15}N-HSQC NMR experiments.

From table 5.10, it can be seen that all compounds are approximately in the same K_D range, between 100 – 500 μM. This is not unusual for fragments, and can be considered reasonable starting points for medicinal chemistry. The highest affinity fragment (CFL02f, K_D 154 μM) has a useful secondary amine for SAR exploration using medicinal chemistry. One of the most common reactions for fragment growing is amide bond formation, due to its simplicity and wide range of available fragments. However, usually, advancing fragments in this way requires detailed, high resolution structural knowledge to determine key interactions. ^{15}N HSQC spectra of each compound titration are shown in appendix 8.4.5.

5.2.7.2. Identification of Ligand Binding Sites by Chemical Shift Mapping

Chemical shift mapping (CSM) is a term used to describe monitoring of chemical shift perturbations and using them to determine the binding site.[238] CSPs are most commonly tracked in ^{13}C and ^{15}N HSQC spectra, however ^{15}N labelling is comparatively cheaper. The most common way of mapping the binding site requires an assigned HSQC spectrum and a detailed structure. CSPs are then mapped to the protein surface by colour coding according to the significance of the chemical shift change. However, usually the entire amino acid is coloured this way, rather than just the backbone amide residue, which can lead to misinterpretation of the essential interactions.[177]

Figure 5.37 shows a histogram illustrating the maximum backbone amide chemical shift changes at saturating amounts of CFL02f ligand. This was achieved for all 6 active fragments (CFL02, CFL02a, CFL02b, CFL02f, CFL17 and CFL17a). There are three thresholds to determine the significance of chemical shift changes: slightly affected (0.04 – 0.10 ppm), moderately affected (0.10 – 0.20 ppm) and strongly affected residues (> 0.20 ppm).

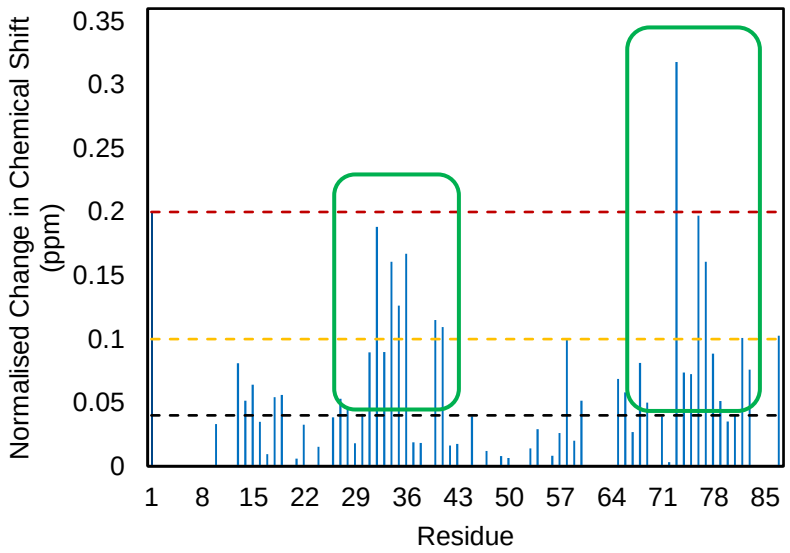


Figure 5.37: CFL02f CSPs per residue. This plot shows a histogram of normalised amide chemical shift changes per residue in the presence of saturating amounts of CFL02f (1.9 mM). Amide chemical shifts were normalised using the equation described in chapter 4. The black dashed line indicates the chemical shift threshold for the slightly affected residues; above the yellow dashed line are moderately affected residues and above the red dashed line are strongly affected residues. The significantly affected residues (yellow and above) are assigned to: E32, K35, R36, Q40, V41, R73, L76. Unassigned residues are not shown. The green boxes show clustering of amino acid residues which is indicative of specific binding.

Using these thresholds, a crystal structure available in the PDB (PDB ID: 3D3B) was colour coded according to the CSPs (this is shown for compound CFL02f in figure 5.38). When a ligand binds to the protein, chemical shift changes can be observed directly at the binding site or by inducing a conformational change in the protein. This can be imagined as an allosteric effect of ligand binding. If a ligand binds to a single site, then any changes far from that location must be due to a conformational change in the backbone of the protein. An encouraging indication from figure 5.37, is that CSPs appear to occur in clusters, which infers a single binding site (green boxes). The yellow residues (figure 5.38) appear to cover a wide area of the protein surface, however, this could be attributed to broad conformational changes.[239] Unfortunately, it is not possible to map all of the changes in chemical shift due to unassigned resonances in the HSQC spectrum. Also, a 180° rotation shows that the rear facing surface also exhibits chemical shift changes (figure 5.38).

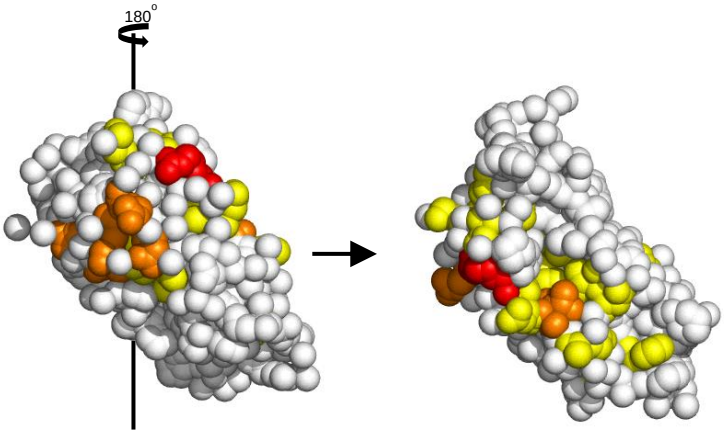


Figure 5.38: Chemical shift map of CFL02f CSPs represented on NusEΔ (PDB 3D3B from *E. coli*). The location of each amino acid affected is colour coded. Slightly affected residues are in yellow, moderately affected residues are in orange and strongly affected residues are in red. All atoms of amino acids whose backbone amide group are affected are coloured.

This process was repeated and a chemical shift map was created for each fragment analogue, these are shown in figure 5.40. Analysing the chemical shift maps for each compound it can be seen that there is a similar clustering of CSPs for each fragment analogue, indicating a similar binding mode. Figure 5.39 illustrates the PPI surface area of NusE/NusB and the affected residues induced by CFL02f. The region highlighted in the red square is the α1 helix of NusE, which appears mostly affected, this is at the PPI surface interface between NusE-NusB, this is a promising result for potential inhibition of this PPI.

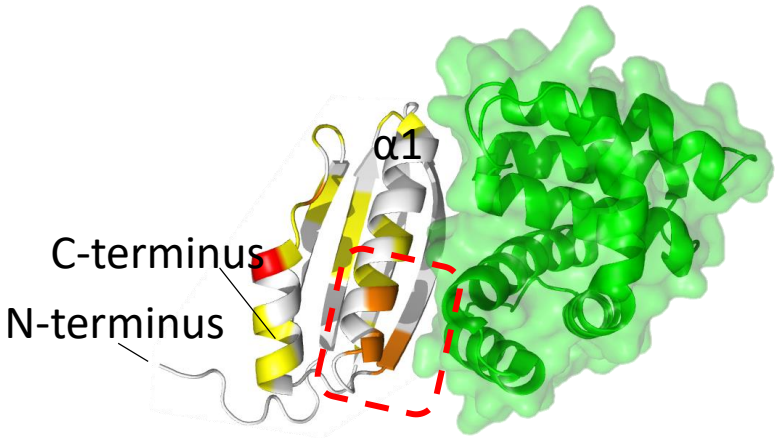


Figure 5.39: NusE-NusB interface. NusB is coloured green and NusE is colour coded according to affected residues (by CFL02f). The red box indicates the most affected region of the α1 helix loop of NusE at the interface with NusB. Yellow residues: slightly affected. Orange residues: moderately affected. Red residues: greatly affected.

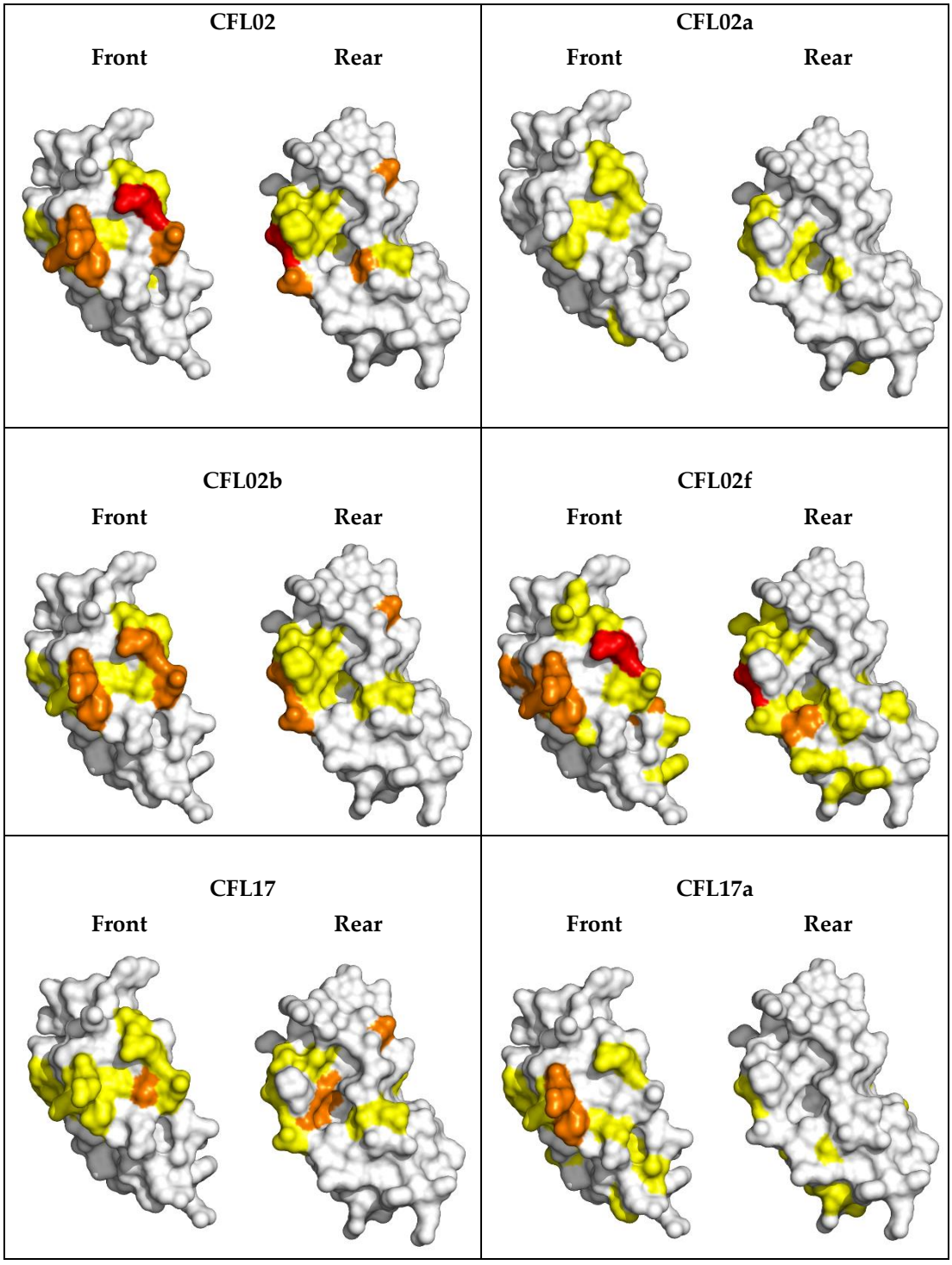


Figure 5.40: Chemical shift mapping of affected residues on NusEΔ (PDB 3D3B) for each compound. Structures were colour coded according to the thresholds: slightly affected residues are in yellow, moderately affected residues are in orange and strongly affected residues are in red.

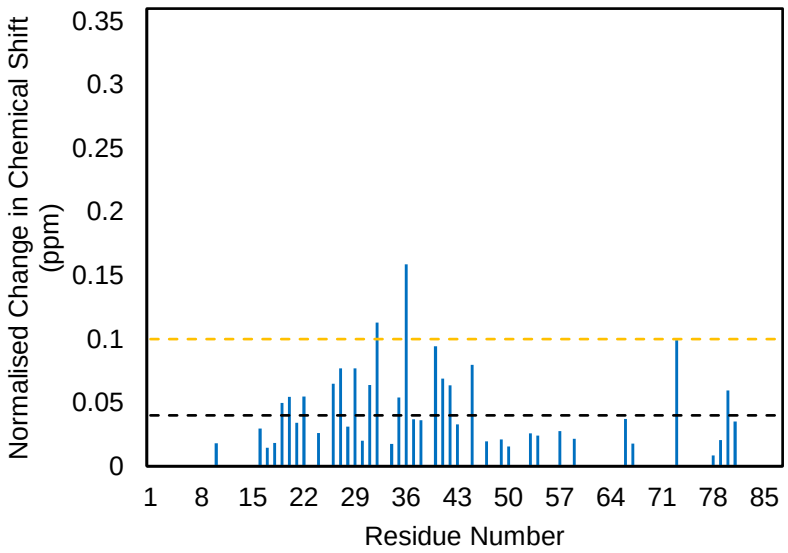


Figure 5.41: CFL17a CSPs per residue. This plot shows a bar chart of normalised amide chemical shift changes per residue in the presence of saturating amounts of CFL02f (1.9 mM). Amide chemical shifts were normalised using the equation described in chapter 4. The black dashed line indicates the chemical shift threshold for the slightly affected residues and above the yellow dashed line are moderately affected residues.

The opposite face of NusE (to NusB) is the PPI surface for the transcription factor NusG (figure 5.42). NusG has been shown to form a complex with NusE and is involved in linking transcription and translation.[240] The ligand binding allosteric effects may therefore also affect the NusG binding capability.

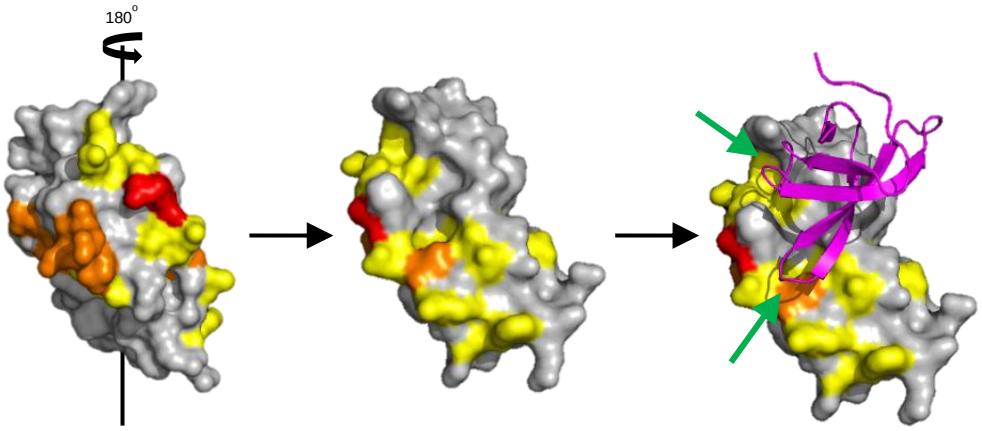


Figure 5.42: The NusE-NusG protein complex (PDB ID NusE: 3D3B alignment with PDB ID NusG: 2KVQ). NusE colour coded with affected residues by CFL02f and NusG in magenta. The affected residues show conformational changes that may affect the NusG binding capability. The PPI contacts of the loop regions on NusG (indicated by the green arrows) are directly affected by the fragment ligand binding.

5.3. Conclusions and future work

In conclusion, two fragments that bind the NusEΔ/NusB PPI were identified from a ^{19}F NMR-based fragment screen. The hit fragments were validated separately and by orthogonal ^{19}F NMR T_2 relaxation experiments. A "SAR by catalogue" approach was used to purchase close analogue fragments and two analogues were found to bind with increased affinity (CFL02f and CFL17a). Microdialysis was adopted as a secondary binding assay to validate the hit compounds from the ^{19}F NMR screen, however, only one compound, CFL02f, was shown to be amenable to this method. The method was validated for determining CFL02f compound affinity for the NusEΔ/NusB protein complex.

Compound	Microdialysis (μM)	dCSP (μM)	^{15}N HSQC NMR
CFL02	-	471	206 ± 16
CFL02a	-	-	455 ± 26
CFL02b	-	-	323 ± 15
CFL02f	171	229 μM	154 ± 3 μM
CFL17	-	590	487 ± 72
CFL17a	-	161	277 ± 65

Table 5.11: A summary of K_D comparison of techniques applied to each compound.

The fragment series for CFL02 and CFL17 were analysed by protein-observed NMR HSQC titration methods for chemical shift mapping and K_D determination. The clustered amide resonances that showed chemical shift changes prove that there is a distinct binding site for each of these fragments. The extension to this work herein is to use CSP-guided docking techniques, such as the HADDOCK approach, to determine the binding mode of fragments and to establish the key interactions.[241,242] Docking studies to identify the binding site would also determine the potential for inhibiting the NusE-NusG interaction pathway.

The dCSP method of K_D determination gives an approximation of the binding affinity and can be a useful method for a preliminary measurement. The K_D from microdialysis for compound CFL02f was in excellent agreement with the ^{15}N-HSQC measured K_D value (table 5.11). The fragments identified here show a promising starting point for development. The most potent fragment, CFL02f, certainly has potential for medicinal chemistry to produce a lead compound. The identified fragments also prove that this PPI is a druggable target and small molecule intervention is certainly a possibility.

To extend this work further, the highest affinity compounds could be used in competition binding assays, such as the FAXS approach (see chapter 4), for further optimisation. These medium affinity ligands are useful as probe molecules and a competition with non-fluorinated analogues is a feasibility. The detailed SAR analysis and cheminformatics capabilities will facilitate in optimising the compounds further. Such small molecular ligands of this novel PPI target will help validate this target as a potential therapeutic strategy.

In summary, the designed in-house CTB libraries produced two novel fragment hits for an under-explored PPI target. This is a hit rate of 2.3 % (Hit rate = (2/87)*100) for the fluorine libraries which is a comparable result to many fragment-based screens.[23]

Chapter 6

Concluding Remarks and Future Outlook

6.1. Conclusions and future work

The origin of disease tends to lie in a complex network of cellular processes mediated by many thousands of biological interactions. Recent technological advances, such as proteomics approaches, to identify these networks has led to the establishment of many PPIs with therapeutic potential.[243] They have long been considered as challenging drug targets due to the large and flat nature of interfaces and also the absence of pockets available for ligand binding. However, the number of successful drugs targeting these interactions is rapidly increasing.[15]

The main aim of the work presented in this thesis was to discover new, novel ligands to difficult, under-explored PPI targets. Owing to the challenges posed by inhibiting this novel class of target new technologies and strategies must be implemented to discover safe and efficacious drug compounds. This was achieved by strategic design and evaluation of peptide and fragment-based libraries for screening by ^{19}F NMR.

6.1.1. Peptides and peptidomimetics

Short peptides are natural ligands involved in PPIs in the cell and therefore the peptide strategy for ligand discovery is well founded. In this project, two parallel approaches to peptide and peptidomimetic ligand discovery were explored, one established and matured by the Auer lab, the other one invented in this work. In the first approach an OBOC peptide library targeted at ATAD2 ligand discovery was successfully designed and synthesised by solid-phase peptide synthesis. This is an example of a target focussed library that can theoretically be used for generic bromodomain ligand discovery. The peptidic tool compound development to investigate the bromodomain containing protein ATAD2, namely H4K5 and H4, were synthesised and modified for detection by NMR spectroscopic methods. However, difficult protein handling conditions and coinciding publication of high affinity small molecules against ATAD2 was a misfortune resulting in abandonment of the project.

The second approach, utilising peptides as screening agents, introduced the concept of 'on-protein peptide growing' (OPPG). This strategy was designed to grow a peptide beginning from a short fluorinated dipeptide, which are generally not used in conventional screening approaches. The use of fluorine as a probe molecule allows the sensitive detection of weak affinity molecules for further optimisation. A successful reverse proof-of-concept study was

achieved with detection of a weak affinity dipeptide peptidomimetic to survivin by ^{19}F NMR. It was also shown that increased affinity towards survivin could be observed using ^{19}F NMR by extension of the dipeptide to the corresponding tripeptide and tetrapeptide. The aim was to use the knowledge acquired via the proof-of-concept peptidomimetic study on survivin to develop a fluorinated dipeptide library using the UbE2L3 target in an initial study. For peptide ligand discovery, two key parameters for optimisation of ligand affinity were applied: the length of the peptide and the nature of the side chains. Therefore, screening a dipeptide library, i.e. the shortest peptide possible, and a library of amino acids, i.e. different side chain containing peptides, tackles both parameters simultaneously. Using solid-phase peptide synthesis (SPPS) a 16 member fluorinated dipeptide library was synthesised for ^{19}F ligand-observed NMR screening. A fluorinated amino acid, a *para*-fluoro-*L*-phenylalanine (pfF), was used as the fluorine containing moiety and these were screened against the UbE2L3 protein. However, initial hit dipeptides showed very low binding affinity and maybe unspecific binding. The most active hit dipeptide was taken forward and a library of 32 C- and N-terminal extended tripeptides were synthesised by SPPS. These were screened for activity against UbE2L3 by ^{19}F NMR however, binding was further reduced.

6.1.2. Fragments

The development of inhibitors for PPI targets has been achieved through a variety of strategies, however, small molecule inhibition remains the preferred method of choice for the pharmaceutical industry. The difficulties of small molecule inhibition of PPIs is that the size of the surface to be covered by the molecule is very large compared to a small active site pocket. However, in fragment-based ligand discovery (FBLD), even though hit compounds are generally weak in affinity, they usually form "high quality" interactions.[244] FBLD has therefore emerged as an attractive option for efficient chemical starting points. One of the first steps in establishing a successful fragment-based approach to ligand discovery is in the design of the library. Fragment libraries can more efficiently probe chemical space and it is this unique trait we wanted to exploit. The aim of this part of the project was to design, construct and evaluate new fluorinated, small fragment libraries and to test them against difficult PPI drug targets. This would ideally result in a more cost effective collection of highly diverse, highly focused fragment libraries, ideally suited for 19F FBS screening in academic labs.

A short analysis of the available commercial fragment libraries allowed us to focus on producing a high quality, affordable and diverse set of fragments. Three separate methods of

fragment library design were implemented in parallel to maximise the quality and diversity of selected fragment sets. The first approach was to select fragments based on the knowledge that they were active in previous biochemical assays. Fragments extracted from the ChEMBL database conforming to a set of criteria were purchased and used to construct the RAFL set of compounds. The second approach, used a common clustering algorithm to explore the entire commercial fragment space. Fragment criteria were defined and these filters were applied to the eMolecules database of commercially available compounds. Fragments that conformed to the selected criteria were clustered according to their 2D molecular similarity. A subjective set of compounds were purchased from cluster sets to maximally represent the fragment space. The third approach to library design introduced a simple metric to ensure maximal diversity when selecting fragments. This was called the A2A score and was used to measure the diversity within commercial and selected fragment sets. This was the core principal of the DDFL library and led to the purchase of fragments that maintained high quality and diversity as a function of cost. The relatively small library sizes described here were created using established methods and shown to be far less expensive than a commercially available library. Regarding the aim of this work, it is likely that the designed, fluorine fragment libraries would not hit PPIs with the same success. Therefore, evaluating the libraries against multiple targets, can assess the quality of diversity and fragments as this eliminates the possibility of frequent hitters and 'bad-acting' compounds.

The designed fluorine fragment libraries (RAFL, CFL and DDFL, collectively known as CTB, chapter 3) were established for use in a ^{19}F NMR-based screen by QC analysis using multiple biochemical methods (HPLC, LCMS and ^{19}F NMR). This rigorous QC analysis was necessary to determine purity and identity and eliminate insoluble or potential aggregating compounds. The assembly of the highest quality ^{19}F fragments formed the basis for the final libraries. The UbE2L3 protein target was screened by ^{19}F NMR, however this fragment screen yielded few hits with high likelihood of mM affinity. Chemical expansion was considered too demanding to tackle in this thesis, and gains from SAR analysis were not thought to present a worthwhile advantage given the limited capacity. It was therefore decided that it would be more important to switch efforts to another PPI target, of different structural and disease context, to exploit and analyse the fragment libraries and gain as much knowledge of their future potential for FBS.

6.1.2.1. NusEΔ/NusB Fragment Screening

The second PPI target introduced was the bacterial antitermination complex NusE/NusB. NusE/NusB is involved at the interface of transcription and translation in prokaryotic species and inhibiting this PPI could lead to new antibiotic therapies. This is a novel, orphan target that at the start of this work, no known inhibitors were currently in existence. From a primary, ^{19}F NMR screen with the CTB fragment libraries two fragments that bind the NusEΔ/NusB PPI were identified from a ^{19}F NMR-based fragment screen. The hit fragments were validated separately and by orthogonal ^{19}F NMR T_2 relaxation experiments. A "SAR by catalogue" approach was used to purchase close analogue fragment series and two analogues were found to bind with increased affinity (CFL02f and CFL17a) to the initial hit compounds. Microdialysis was adopted as a secondary binding assay to validate the hit compounds from the ^{19}F NMR screen, however, only one compound, CFL02f, was shown to be amenable to this method. The method was validated for determining CFL02f compound affinity for the NusEΔ/NusB protein complex. The fragment series analogues for initial hits CFL02 and CFL17 were analysed by protein-observed NMR [^{1}H, ^{15}N]-HSQC titration methods for chemical shift mapping and K_D determination. The clustered amide resonances that showed chemical shift changes prove that there is a distinct binding site for each of these fragments. As outlined previously, the NusG binding site, to NusE, is located on the opposite face to the NusB site (figure 6.1). To ascertain whether the compound (CFL02f) affects the NusG binding capability, a competition binding experiment using ^{15}N-labelled NusE in a similar HSQC assay would determine the effect on NusG. The K_D of the NusE/NusB/CFL02f complex is known (~150 μM), therefore a titration with NusG would allow on to calculate the K_D of the NusG interaction if they were to bind at the same site. In a competitive binding titration experiment, this would result in a cubic algebraic equation, which would be difficult to solve by non-linear least square regression, as in the case of a simple quadratic.[245] Also, another experiment to determine dissociation of the NusE/NusB/NusG complex would be to analyse the transverse relaxation rates of the labelled NusG protein in the presence and absence of inhibitor CFL02f. The formation of the large NusE/NusB/NusG complex would cause increased transverse relaxation times, leading to broadening (and even disappearance) of NMR resonances. Addition of a compound that dissociates the complex to the NusE/NusB heterodimer and NusG protein would restore the HSQC resonances of NusG, whereas a weak or no binder would have no effect on the HSQC spectrum of the complex. Furthermore, this PPI directly binds ssRNA, therefore it is valid to compare the CSP map with the RNA binding track of the

complex. This is shown in figure 6.1 (B and C), where it can be seen that the RNA binding track is located away from fragment binding site and would not appear to be affected by small conformational changes. The ssRNA, shown in figure 6.1 (B and C), is a model complex based on the alignment of two crystal structures from separate organisms (PDB 3D3B and 3R2D).

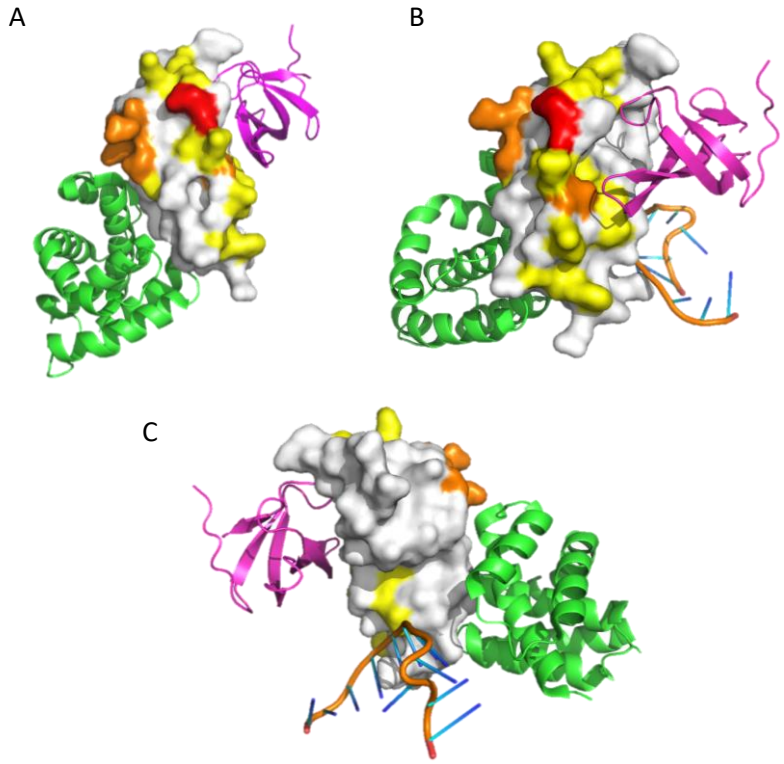


Figure 6.1. Model structure of the NusEΔ/NusB/NusG/RNA complex based on alignment of three separate crystal structures (PDB 3D3B, 2KVQ and 3R2D). NusB is the green ribbon, NusG is the magenta ribbon, NusEΔ is the white surface structure, colour coded according to CFL02f affected residues and the ssRNA is orange/blue. **A.** The NusEΔ/NusB (PDB ID 3D3B) is aligned to NusEΔ/NusG-CTD (PDB ID 2KVQ). Both structures are from *E. coli*. **B**. The rear face of NusEΔ/NusB/NusG complex. Also shown is an ssRNA aligned from PDB ID 3R2D (Organism *Aquifex aeolicus*). **C.** Another view of the RNA binding track of the NusEΔ/NusB/NusG-CTD heterotrimeric complex.

Also, it is worth examining the specificity of these fragments for the NusE/NusB binding interaction between bacterial species. ConSurf is an online bioinformatics tool that allows one to assess the evolutionary conservation rate of amino acids in a protein based on the phylogenetic relations between homologous sequences.[246–250] The degree with which an amino acid in a protein is conserved is greatly dependent on its functional and structural significance. Therefore, conservation analyses of this type across species can reveal the importance of the amino acid to its function. Figure 6.2 shows the degree to which the amino acids in NusE are conserved across species and this is compared to the affected residues by compound CFL02f

binding. As can be seen from the figure, the orange patch of moderately affected residues (E32, K35, R36, Q40 and V41) are shown to be variable across species, indicating these residues are not important to the binding interaction.

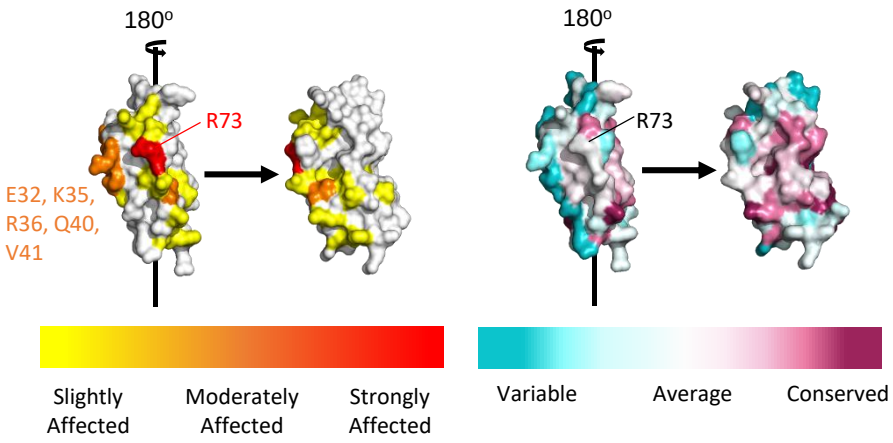


Figure 6.2: Comparison of NusE CSP map (left) and ConSurf results (right).

The most common approach to targeting PPI complexes is to develop a compound that inhibits or completely dissociates the interaction. Therefore, further experiments to determine the effect on the NusEΔ/NusB complex are required. Although there is no evidence for dissociation of the NusE/NusB PPI complex by these small fragments, further methods to elucidate the mode of action might include fluorescence-based assays.

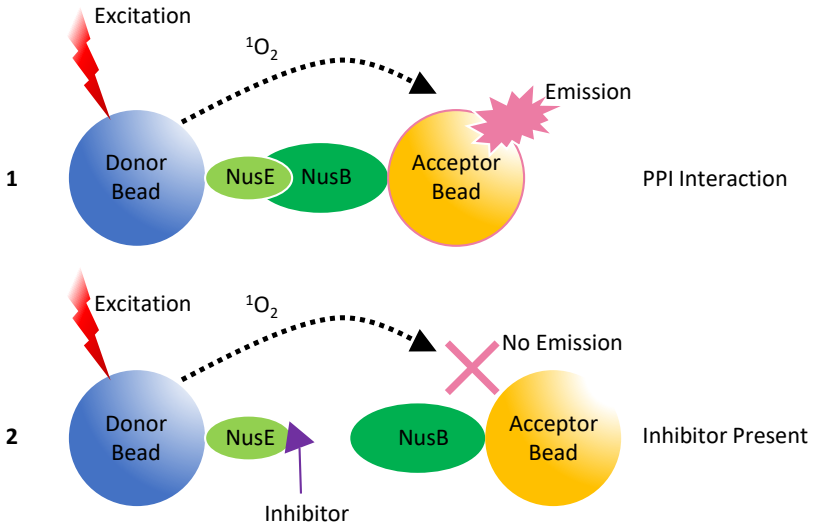


Figure 6.3: Schematic of the AlphaScreen™. 1. Excitation of the donor bead causes release of singlet oxygen. In the presence of a biological PPI, the acceptor bead is in close proximity and emission occurs. 2. In the presence of an inhibitor, the acceptor bead is too far for singlet oxygen to travel, therefore, a decrease in fluorescence is observed.

For example, the AlphaScreen™ (Amplified Luminescent Proximity Homogeneous Assay, PerkinElmer) is a fluorescent assay based on the use of donor and acceptor beads functionalised with the target proteins (figure 6.3). In the presence of a biological interaction, the singlet state oxygen produced by excitation of the donor bead reacts with fluorophores in the acceptor bead to generate a fluorescent response. However, a decrease in signal would be observed in the presence of an inhibitor of this interaction. This could be used to identify if, at all, the NusE/NusB PPI is disrupted on binding of the hit fragments. A further experiment to determine the binding site would be to utilise site-directed mutatgenesis to change one of the affected residues and observe the binding effect of the ligand. For example, the most strongly affected residues is R73, therefore incorporating an alanine in this position and measuring the K_D would provide information on the binding site. Also, as shown previously, the fragment CFL02f appears to reduce the growth of *E. coli* cells at high concentrations (section 5.2.6). To examine if the fragment disrupts the NusE/NusB PPI, one could use a strain that overexpresses one component of the complex. To compare both strains, using a constant inhibitor concentration, bacterial growth may be restored in the overexpression strain.

The fragments identified here show a promising starting point for development. The most potent fragment, CFL02f, certainly has potential for medicinal chemistry to produce a lead compound. The identified fragments also prove that this PPI is a druggable target and small molecule intervention is certainly a possibility.

6.2. Summary and Future Outlook

The discovery of ligands for PPI targets is a challenging area. The methods employed in this thesis aimed to maximise the value of established techniques for use in an academic setting. ^{19}F NMR is acknowledged as a valuable primary screening method in industrial and academic drug discovery and the peptide and fragment approaches to ligand discovery both deserve merit. Establishment of a PPI screening platform using small libraries of fluorinated compounds, either peptides or fragments, for novel ligand discovery was shown to be well suited and easily implemented in an academic lab.

The peptide and peptidomimetic approach to ligand discovery offers the advantage that straightforward synthesis methods can lead to many thousands of compounds in a short timescale. The OPPG approach described here is novel in that it uses 'fragment-sized' compounds (dipeptides) and a highly sensitive probe to detect binding. Simple C- or N-

terminal peptidic extension can quickly lead to high affinity peptides for applications in basic and translational science, particularly for the set-up of competitive HTS, for structural biology and for drug delivery. To develop the OPPG approach further, additional E2 enzymes could be screened with the first fluorine dipeptide library. This would provide a measure of the druggability of this family of proteins and, if successful, would offer vital information for the pharma industry. It would also indicate how represented pockets able to accommodate the core *m*CF3-Phe amino acid derivative are on E2 enzymes. However, this approach can easily be extended to using different fluorinated amino acids as the chemical probe and non-natural amino acids in the 'N' and 'C' positions. This novel approach to 'on-protein peptide growing' can theoretically be used against any target. The relative ease of peptide synthesis, coupled with the assay sensitivity is a useful novel approach to peptide ligand discovery.

The fluorinated fragment libraries were designed to be affordable, high quality and highly diverse and the results herein are evidence of this. The A2A metric described here is a simple measure of diversity that can be exploited by other, small academic groups looking to construct small, focussed, affordable fragment libraries. It can also be further implemented to expand the chemical space of existing fragment libraries for efficient drug discovery. The designed in-house CTB fragment libraries produced two novel fragment hits for an under-explored PPI target, NusE/NusB. This is a hit rate of 2.3 % for the fluorine libraries which is a comparable result to many fragment-based screens.[23] Structural data for PPI targets is almost always mandatory[251], therefore, the extension to this work is to use CSP-guided docking techniques, such as the HADDOCK approach, to determine the binding mode of fragments and to establish the key interactions.[241,242] Docking studies to identify the binding site would also determine the potential for inhibiting the NusE/NusG interaction pathway. To extend this work further, these medium affinity ligands are useful as probe molecules and a competition binding assay with non-fluorinated analogues is a feasibility, such as the FAXS approach (described by Dalvit *et al*, see chapter 4).The detailed SAR analysis and cheminformatics capabilities of the Auer group will facilitate in optimising the compounds further. Such small molecular ligands of this novel PPI target will help validate this target as a potential therapeutic strategy.

Although fluorine containing chemical space only covers a small fraction of overall chemical space, ^{19}F FBS as a primary NMR screening method to detect medium affinity ligands allows one to expand the chemical space for competition-based assays with non-fluorinated

analogues. Ideally, one would have a small, diverse set of fluorinated fragments, and each fragment would have a sub-library of non-fluorinated analogues (i.e. SAR-related sub-library). Discovery of an initial fluorinated fragment hit, the SAR sub-library could be used in competition-based assays. ^{19}F NMR may hold the key for more efficient ligand discovery of PPI targets.

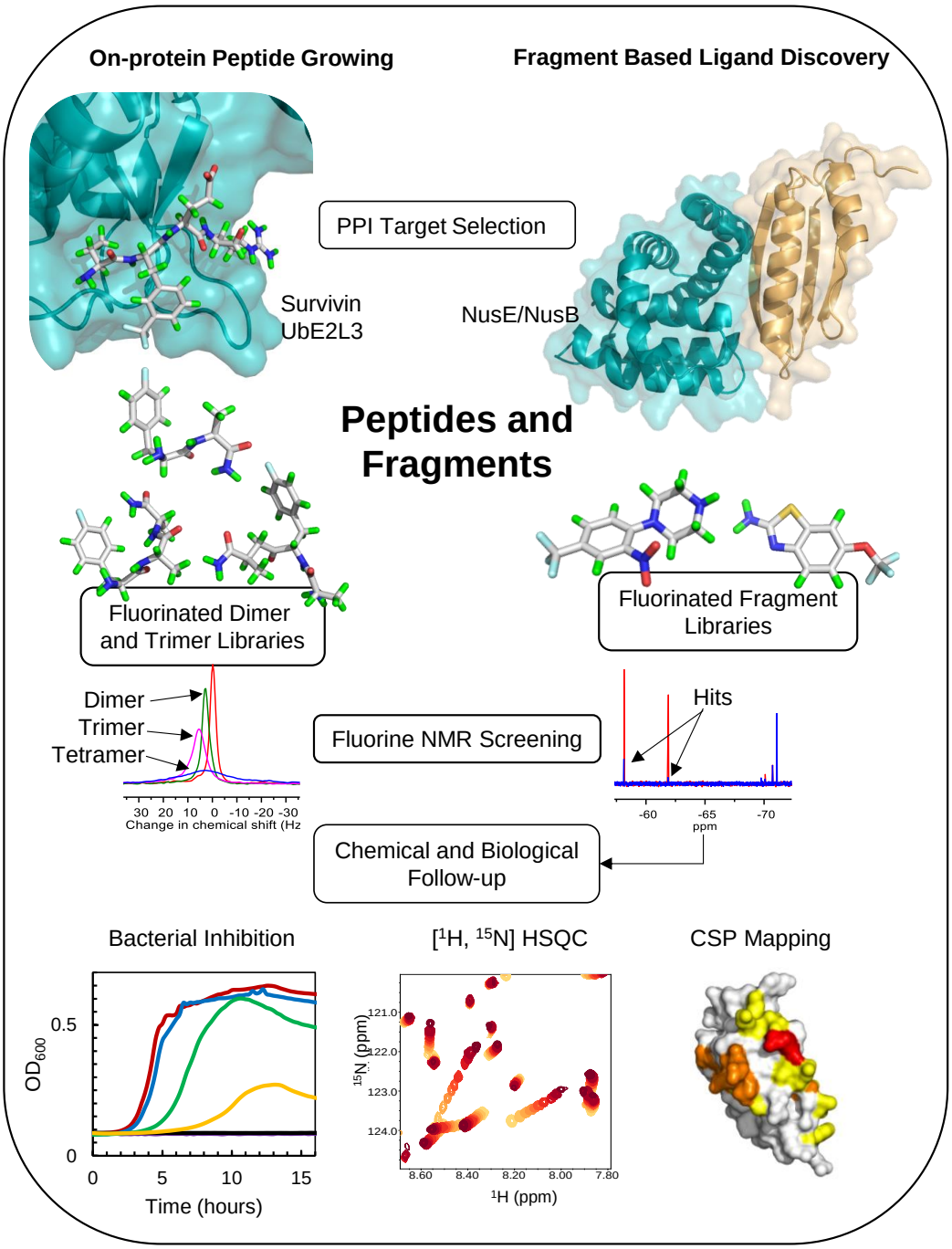


Figure 6.1: Complete scheme to successful ligand discovery methods applied in this thesis.

Chapter 7

Materials and Methods

This chapter details the experimental methods and reagents used for each chapter. The methods are separated into specific chapters. All commercially available reagents were purchased from Sigma Aldrich or VWR unless otherwise stated and used as received after quality control by HPLC and LCMS.

7.1. General Analytical and Synthetic Methods

7.1.1. LCMS Analysis

LCMS analyses (p were performed on an Agilent 1100 micro-HPLC system, equipped with a capillary pump (G1376A), a micro-autosampler (G1377A) and a diode-array detector (DAD; G1315B) and combined with a Finnigan Deka CP Plus ESI-MS detector. The system was run with a spray voltage of 5 kV, capillary temperature of 275 °C and a capillary voltage of 15 V in positive single-ion MS mode. The HPLC system consisted of a Zorbax-SB-C18 column (0.5 x 35 mm) with a 3.5 μm particle size and the solvent system used was: A (H_2O, 0.1 % TFA) and B (MeCN, 0.1 % TFA). The standard gradient used for compound analysis was: 1.5 min 0 % B, 14 mins 0 – 100 % B, 15 mins 100 % B at a flow rate of 60 $\mu l.min^{-1}$. HPLC-ESI-MS data were processed using the Xcalibur software package (version 2.0, Thermo Electron Corporation, MA, USA).

7.1.2. Analytical HPLC

HPLC compound analyses were performed on an Agilent 1100 series HPLC system, equipped with a quaternary pump (G1311A), a degasser (G1322A), a well-plate autosampler (G1367A), a fluorescence detector (FLD; G1321A) and a DAD (G1315B). Solvent system: A (H_2O, 0.1% TFA) and B (MeCN, 0.1% TFA).

Method 1: Column: Agilent ZORBAX SB-C18 (4.6 x 150 mm) with a 3.5 μm particle size diameter and 80 μm pore size. Gradient: 0 - 5 min, 5 % B; 5 – 24 min, 5 - 100 % B, 24 – 26 min, 100 % B; 26 - 27 min, 100 - 5 % B; 27 – 30 min, 5 % B with a flow rate of 0.8 $ml.min^{-1}$.

Method 2: Column: Vydac C4 (RP-C4; 2 x 125 mm). Gradient: 0 - 5 min, 5 % B; 5 – 24 min, 5 - 100 % B, 24 – 26 min, 100 % B; 26 - 27 min, 100 - 5 % B; 27 – 30 min, 5 % B with a flow rate of 0.8 $ml.min^{-1}$.

Method 3: Column: Vydac C4 (RP-C4; 2 x 125 mm). Gradient: 0 - 5 min, 5 % B; 5 – 35 min, 5 - 95 % B; 35 – 40 min, 95 % B; 40 - 45 min, 95 - 5 % B; 45 – 50 min, 5 % B with a flow rate of 0.8 $ml.min^{-1}$.

The retention time (t_R) for analytes is given in minutes and purity is calculated using the peak integral. DAD HPLC trace measurements were taken at 210, 220, 254 and 280 nm. For TMR

labelled peptides, a fluorescence trace was measured using an excitation wavelength (λ_{ex}) of 555 nm and emission wavelength (λ_{em}) of 575 nm on the FLD. For Cy5-labelled protein, a fluorescence trace was measured using an excitation wavelength (λ_{ex}) of 650 nm and emission (λ_{em}) wavelength of 670 nm on the FLD.

7.1.3. NMR Spectroscopy

The University of Edinburgh, School of Chemistry has a range of liquid state Bruker Avance III NMR spectrometers with a variety of probes for characterisation of compounds. NMR experiments were performed, where stated, using spectrometers: Deuterated solvents used for compound analyses were MeOD and DMSO-d_6 where stated.

- Bruker Avance III 400 MHz spectrometer (AVA 400), equipped with BBFO$^+$ room temperature probe and two channels, ^{1}H and ^{19}F (optimised)
- Bruker Avance III 500 MHz (PRO 500) spectrometer equipped with Prodigy CryoProbe and two channels ^{1}H/^{19}F and X nuclei (optimised)
- Bruker Avance III 500 MHz (AVA 500) spectrometer equipped with DCH Probe and two channels: ^{1}H and ^{13}C (optimised)
- Bruker 800 MHz spectrometer (AVA 800) equipped with TCI CryoProbe and three channels: ^{1}H (optimised), ^{13}C and ^{15}N.

Two dimensional [^{1}H, ^{15}N]-HSQC NMR for binding site mapping and K_D determination of compounds against the NusEΔ/NusB protein was completed at the University of Bayreuth, Germany, using a Bruker Aeon 1 GHz (AEON 1000) spectrometer.

Chemical shifts (δ) are quoted in parts per million (ppm) downfield of tetramethylsilane, using residual protonated solvent as internal standard ($CDCl_3$ at 7.27 ppm or acetonitrile-*d*3 at 1.94 ppm). Abbreviations used in the description of resonances are: s (singlet), d (doublet), t (triplet), q, (quartet), app (apparent), br (broad). Proton-decoupled 13C NMR spectra were recorded on a Bruker AV500 (125.8 MHz) spectrometer or a Bruker AVA400 (100.6 MHz) spectrometer. Chemical shifts (δ) are quoted in parts per million (ppm) downfield of tetramethylsilane, using deuterated solvent as internal standard ($CDCl_3$ at 77.0 ppm or acetonitrile-d3 at 118.7 ppm and 1.4 ppm).

7.1.3.1. ^{19}F NMR Parameters

Fluorine NMR peptide and fragment screening and QC analyses were recorded at 300 K on a Bruker Avance 400 MHz spectrometer with BBFO^{+} room temperature probe for direct ^{19}F detection. One dimensional ^{19}F spectra were acquired for each sample with ^{1}H decoupling according to the parameters described in table 7.1. Sample preparation for peptide and fragment pools are given in section 7.2.12 and 7.3.1 respectively.

Experiment	SFO1 (MHz)	SW (Hz)	TD	NS	AQ (s)
^{19}F NMR	376.47	111111.1	262144	256	1.18

Table 7.1: ^{19}F NMR spectroscopy parameters. SFO1: Irradiation frequency channel 1; SW; Spectral width; TD: Time domain; NS; Number of scans; AQ: Acquisition time.

7.1.3.2. [^{1}H, ^{15}N]-HSQC NMR Experiments

^{15}N-HSQC experiments were recorded at 300 K on Bruker Avance 800 MHz for the ATAD2 studies and Bruker Aeon 1 GHz NMR spectrometer for NusEΔ/NusB. Both spectrometers are equipped with cryogenically cooled probes for optimum sensitivity. Sample preparation is given in the respective sections.

7.1.4. Software

All PDB protein structures are analysed and illustrated using PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Processing of 1D NMR data and analysis of chemical shifts and intensities was performed using in-house routines and visualised using MestReNova NMR software package (v10.0.2-15465, Mestrelab Research, S.L.). Processing of 2D NMR spectra was performed using NMRViewJ (v. 9.2.0-b11, One Moon Scientific Inc.). Charts and graphs were illustrated using Microsoft Excel (2013) and Origin 9.0 (OriginLab Corporation) where stated. K_D simulations and fittings were performed in GraFit 7.0.3 (Erithacus Software Ltd.). Chemical structures were drawn using ChemDraw (v 14.0.0.117, CambridgeSoft Corporation, PerkinElmer, Inc).

7.1.5. General Fmoc Solid-phase Peptide Synthesis (SPPS) Methods

TentaGel™ rink amide resin (0.26 mmol.g^{-1}, 90 μm, TG-RAM) was supplied by Rapp Polymer. Amino acids and all other reagents required for SPPS were supplied by Novabiochem, Chem

Impex and Sigma Aldrich. There are three steps to Fmoc-based SPPS: 1) amino acid coupling, 2) Fmoc deprotection and 3) cleavage from the resin.

Resin preparation

TentaGel™ (TG) resin (0.23 mmol.g^{-1} 90 μm or 0.3 mmol.g^{-1} 140 μm resin, Polymer) is weighed out for the desired yield of peptide. The resin is added to an SPE fitted with a frit and swollen in DMF (1 ml/100 mg resin) for 30 minutes.

Amino Acid Coupling

To a solution of amino acid (3 eq.) and HATU (2.5 eq.) in DMF (1 ml/100 mg resin) is added DIPEA (6 eq.). The solution is then transferred to an SPE containing the resin fitted with a frit. The SPE is capped and shaken for 20 – 30 minutes, then the solution drained and the coupling reaction repeated. The resin is then washed extensively with DMF, H_2O, MeOH and finally DCM to remove any traces of coupling agents. The resin is either dried overnight *in vacuo* in preparation for cleavage, or swollen again in DMF for the next amino acid.

Fmoc Deprotection

To the Fmoc-protected amino acid resin, a solution of 20 % piperidine in DMF (1 ml/100 mg resin) is added. The mixture is shaken for 15 minutes then the resin drained and rinsed with 10 column volumes of DMF. The deprotection reaction is repeated once again for a further 10 minutes.

TNBS test for primary amines

The TNBS test is a sensitive test to monitor the completion of the coupling reaction.[252] It can only be used for the detection of primary amines which are visualised by an intense orange colour under a microscope. Only the beads with a primary amino group will turn orange/red and the intensity of the colour does not depend on the nature of the N-terminal amino acid.

A small sample of resin is placed into a glass vial and suspended in 200 μl of 10 % DIPEA in DMF. A small drop of 5 % trinitrobenzenesulfonic acid (TNBS) in water is added to the resin and the mixture shaken. The vial is then viewed under a microscope to determine the coupling reaction or the Fmoc deprotection efficiency. If any orange beads are present, the coupling reaction is repeated. If non-coloured beads are observed the coupling is repeated.

Standard cleavage from rink resin

The resin is dried overnight *in vacuo* to remove any traces of DCM from the previous washing steps. A solution of 2.5 % H_2O, 2.5 % TIS in TFA (1 ml/100 mg resin) is prepared and added directly to the resin and shaken for 3 hours. The resulting solution is drained into a glass vial and volatiles removed under reduced pressure (Savant SpeedVac ® Plus SC110A, equipped with Refrigerated Vapor Trap, RVT400) . Following this, ice-cold diethyl ether is added to the residue and sonicated to break up any large clumps. The solution is centrifuged briefly and the supernatant is discarded. This process is repeated 3 times and the remaining solid is dried overnight *in vacuo*.

7.2. Experimental Section for Chapter 2

7.2.1. ATAD2 Expression and Purification

Reagents and buffers for ATAD2 expression and purification are indicated in table 7.2

Chemically Competent Cell Preparation

A starter culture (5 ml) of *E. coli* expression strain Rosetta (DE3)pLysS was grown overnight at 37 °C. The cells were harvested by centrifugation at 3000 *g* for 15 minutes at 4 °C. The cell pellets were then resuspended in 50 ml of ice cold $CaCl_2$ (30 mM) by vigorous vortexing. The cells were centrifuged at 3000 *g* for 15 minutes at 4 °C. The supernatant was discarded and cells resuspended in 50 ml ice cold $CaCl_2$ (30 mM). The cells were left on ice for 20 minutes then harvested by centrifugation at 3000 *g* for 15 minutes at 4 °C. The supernatant was discarded, and the pellet carefully resuspended in 50 ml ice cold 85 mM $CaCl_2$ with 15 % v/v glycerol. The cells were left on ice for 20 minutes then harvested by centrifugation at 3000 *g* for 15 minutes and 4 °C. Again, the supernatant was discarded, and the cells resuspended in 2 ml ice cold 85 mM $CaCl_2$ with 15 % v/v glycerol. The cells were then divided into 50 μl aliquots in pre-chilled micro test tubes and flash frozen in liquid nitrogen. The competent cell fractions were stored at -80 °C.

ATAD2 Buffers	Components
Binding buffer	50 mM HEPES, 500 mM NaCl, 5 mM imidazole, 0.5 mM TCEP, pH 7.5.
Wash buffer	50 mM HEPES, 500 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, pH 7.5.
M9 minimal stock solution (5X)	30 g.L^{-1} Na_2HPO_4, 15 g.L^{-1} KH_2PO_4, 2.5 g.L^{-1} NaCl
TS2-Trace element stock solution	100 mg.L^{-1} $ZnSO_4.7H_2O$, 30 mg.L^{-1} $MnCl_2.4H_2O$, 300 mg.L^{-1} H_3BO_3, 200 mg.L^{-1} $CoCl_2.6H_2O$, 10 mg.L^{-1} $CuCl_2.2H_2O$, 20 mg.L^{-1} $NiCl_2.6H_2O$, 900 mg.L^{-1} $Na_2MoO_4.2H_2O$, 20 mg.L^{-1} Na_2SeO_3
M9-minimial medium	1X M9-stock solution, 1.5 g.L^{-1} $(^{15}NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.2 % (v/v) TS2-trace element solution, 10 μM Fe(III) citrate, 100 μM $CaCl_2$, 0.4 % (w/v) glucose, 1X MEM-vitamins.

Table 7.2: Reagents and components of buffers used for the purification of ATAD2 protein.

Transformation

A 1 μl solution of prepared ATAD2 plasmid (supplied by SGC, Oxford, 10 ng.μl^{-1}) was added to 50 μl of chemically competent cells. The cell/plasmid solution was mixed gently by flicking and the mixture left on ice for 30 minutes. The micro test tube containing the cell/plasmid mixture was then submerged in a water bath at 42 °C for 45 seconds exactly then immediately returned to ice and left for 5 minutes. The transformed cell solution was then added to 25 ml of pre-warmed LB broth and incubated for 1 hour at 37 °C. After 1 hour, Kanamycin was added to a final concentration of 50 μg.ml^{-1} and the solution grown overnight at 37 °C. The next day a 25 % glycerol stock was made by adding 1 ml of 50 % glycerol to 1 ml of culture and the stock stored at – 80 °C.

Expression and Purification

A small aliquot of the ATAD2 glycerol stock was inoculated into 5 ml LB broth containing 50 μg.ml^{-1} Kanamycin. The culture was grown overnight at 37 °C with agitation (180 rpm). A 1 ml aliquot from the overnight culture was inoculated into 1 L of LB broth. The culture was incubated at 37 °C until the OD_{600} reached 0.5 AU and then the temperature adjusted to 18 °C. Expression was induced by adding IPTG to a final concentration of 0.5 mM and the culture incubated overnight at 18 °C with shaking (180 rpm).

The 1 L culture was harvested by centrifugation (Sorvall RC 5C Plus) at 3500 *g* for 15 minutes and 4 °C, the supernatant was discarded. The pellet was resuspended in 50 ml of binding buffer and mixed vigorously by vortexing. The cell culture was incubated on ice for 30 minutes prior to sonication. The cells were kept on ice and sonicated for 1 minute with 10 second pulses (at 50 % amplitude) with 10 second intervals. The lysate was incubated at 4 °C for 5 minutes with gentle rotation and then sonicated again as above. The lysate was cleared by centrifugation at 4 °C and 4000 rpm (3500 *g*) for 1 hour.

For protein purification of ATAD2 see section 7.2.4.

7.2.2. ^{15}N Labelling and Expression of ATAD2

E. coli (BL21) expression strain transformed with the ATAD2 containing plasmid (~20 μl) was plated onto LB agar plates containing Kanamycin (50 μg.ml^{-1}) and grown overnight at 37 °C in an incubator. A single colony was picked and inoculated into 5 ml LB broth containing Kanamycin (50 μg.ml^{-1}) and incubated overnight at 37 °C with agitation at 180 rpm. This 5 ml culture was then used to inoculate a 1 litre culture of M9-minimal media (containing $(^{15}NH_4)_2SO_4$ (Sigma Aldrich 299286) as the sole nitrogen source) which was incubated at 37 °C until the OD_{600} reached 0.5 AU (with agitation at 180 rpm). Induction of expression was achieved by supplementing the media with a final concentration of 0.5 mM IPTG and the culture incubated overnight at 18 °C with agitation (180 rpm).

Cell harvest and lysis

The cell culture was pelleted in a centrifuge (Sorvall RC 5C Plus) at 4 °C and 4000 rpm (3010 *g*) for 30 minutes. The supernatant was removed, and the cell pellet resuspended in 40 ml of ice cold PBS (by vortexing) and then centrifuged again at 4 °C and 4000 rpm (3010 *g*) for 30 minutes. The cell pellet was then stored before purification at – 20 °C. The frozen cell pellet was left to thaw at room temperature for approximately 1 hour prior to lysis. The cell pellet was resuspended (by vortexing) in 50 ml of binding buffer and incubated on ice for 30 minutes prior to sonication. The cells were kept on ice and sonicated for 1 minute with 10 second pulses (at 50 % amplitude) with 10 second intervals. The lysate was incubated at 4 °C for 5 minutes with gentle rotation and then sonicated again as above. The lysate was cleared by centrifugation at 4 °C and 4000 rpm for 60 minutes.

For purification of ^{15}N-ATAD2 see section 7.2.4

7.2.3. Ubiquitin Conjugating Enzyme L3 (UbE2L3) Expression and Purification

Reagents and buffers for UbEL3 expression and purification are indicated in table 7.3

UbE2L3 Buffers	Components
Minimal buffer	50 mM sodium phosphate, 500 mM NaCl, 5 mM imidazole, 10% glycerol, pH 7.5
Lysis buffer	1X minimal buffer, 5 mM β-mercaptoethanol, protease inhibitor cocktail (1 tablet.50 ml^{-1} EDTA free, Roche), lysozyme (from chicken egg white, 43000 $units.mg^{-1}$ solid, Sigma Aldrich), benzonase nuclease (> 250 $units.\mu l^{-1}$, Sigma Aldrich), pH 7.5
Wash buffer	1X minimal buffer, 15 mM imidazole, 5 mM β-mercaptoethanol, pH 7.5
Elution buffer	1X minimal buffer, 250 mM imidazole, pH 7.5
Dialysis buffer	20 mM HEPES, 100 mM NaCl, pH 7.5

Table 7.3: Components of buffers used in the purification of UbE2L3.

Protein Expression

E. coli BL21(DE3) bacterial expression strain transformed with the E2 sequence containing plasmid (supplied by Joanna Koszela, pET28a) (10 μl) was plated onto LB agar plates containing Kanamycin (50 $\mu g.ml^{-1}$) and grown overnight at 37 °C in an incubator. A single colony was picked and inoculated into 10 ml LB broth containing Kanamycin (50 $\mu g.ml^{-1}$) and incubated overnight at 37 °C with agitation at 180 rpm. The next day, 10 ml of culture was inoculated into 1 L of fresh LB broth supplemented with Kanamycin (50 $\mu g.ml^{-1}$) and grown at 37 °C with agitation (180 rpm) until the culture reached an OD_{600} of 0.3 - 0.5 AU. Induction of expression was achieved by supplementing the media with a final concentration of 0.5 mM IPTG and the culture incubated overnight at 20 °C with shaking (180 rpm).

Cell harvest and lysis

The overnight culture was spun down in a centrifuge (Sorvall RC 5C Plus) at 4 °C and 4000 rpm (3010 *g*) for 30 minutes. The supernatant was removed, and the cell pellet resuspended in 40 ml of ice cold PBS (by vortexing) and then centrifuged again at 4 °C and 4000 rpm (3010 *g*) for 30 minutes. The cell pellet was then stored before purification at – 20 °C. The frozen cell pellet was left to thaw at room temperature for approximately 1 hour prior to lysis. The cell pellet was resuspended (by vortexing) in 30 ml of lysis buffer, supplemented with protease

inhibitor cocktail, 30 mg lysozyme and 0.5 µl Benzonase® nuclease (Sigma Aldrich, E1014, 250 units.µl^{-1}). The cell suspension was incubated on ice for 30 minutes prior to sonication. The cells were kept on ice and sonicated for 1 minute with 10 second pulses (at 50 % amplitude) with 10 second intervals. The lysate was incubated at 4 °C for 5 minutes with gentle rotation and then sonicated again as above. The lysate was cleared by centrifugation at 4 °C and 4000 rpm for 30 minutes, the supernatant removed, then centrifuged at 10000 rpm for 15 minutes.

For protein purification of UbE2L3 see section 7.2.4.

7.2.4. General Protein Purification Protocol

The proteins (ATAD2, [^{15}N]-ATAD2 and UbE2L3) were expressed with a 6 x His-tag at the *C*-terminus for purification by Nickel NTA affinity chromatography. Nickel (II)-NTA agarose beads (1 ml, 50 % slurry) (Qiagen, 30250) were washed with wash buffer (3 x 15 ml) to remove excess ethanol and the cleared lysate incubated with the resin at 4 °C for 2 hours with gentle rotation. The bead mixture was poured over a 20 ml column (BioRad) and the beads left to settle at the bottom. The flow through was collected by gravity flow for analysis by SDS-PAGE and the column washed twice with 20 times bed volume of lysis buffer containing 20 mM imidazole (and preventing the beads from drying out). The 6xHis-tagged proteins were eluted in 8 x 2 ml fractions using a stepwise gradient of 50 – 250 mM imidazole (50, 75, 100, 150, 200, 250, 250, 250 mM). The fractions were run on SDS-PAGE to assess purity. Fractions containing protein were pooled and dialysed (3500 MWCO, Thermo Scientific Slide-A-Lyzer dialysis cassette) into 1 L of dialysis buffer (see table 7.3) for 1 hour at 4 °C with gentle stirring. The cassette was then transferred into 4 L of fresh dialysis buffer and left overnight at 4 °C, with stirring. The following day, the dialysed protein solution was concentrated in a 20 ml protein concentrator spin column (Thermo Scientific Pierce, PES MWCO 10,000) using a bench top centrifuge set at 4000 rpm and 4 °C until the volume was reduced to 500 µl. After concentration determination (see section 7.2.5), to UbE2L3 DTT was added to a final concentration of 1 mM (UbE2L3). The pooled protein fractions were then aliquoted into 30 µl fractions and flash frozen in liquid nitrogen and stored at - 80 °C.

The proteins were analysed by HPLC (Method 2 for UbE2L3 and Method 3 for ATAD2) and SDS PAGE for purity and the corresponding gel images and chromatograms are given in appendix 8.1.

7.2.5. Concentration determination by UV-Vis Spectroscopy

The protein concentration was determined by making a 1:10 dilution of purified protein in elution buffer and then measured using a NanoDrop 1000 spectrophotometer (PEQLAB biotechnologie GmBH). The volume of the sample is 2 μl and the measurement was recorded three times and averaged for accurate concentration determination. The protein extinction coefficients

7.2.6. SDS-PAGE

SDS-PAGE analysis was carried out according to standard methods and procedures. The gels used were precast polyacrylamide Bolt 12 % Bis-Tris Plus Gels (1.0 mm, 17 well, Invitrogen, Thermo Fisher Scientific) and run using MES SDS Running buffer (20x, Novex, Life Technologies). The gel was run using 200 V Bio-Rad Power Pac 3000 for 30 min, the loading buffer was Bolt™ LDS sample buffer (4x, Novex, Life Technologies), molecular weight ladder was SeeBlue Plus2 Prestained (Invitrogen, Thermo Fisher Scientific). Standard procedure involves mixing sample with loading buffer under reducing conditions (1 M DTT) and heating the sample for 10 minutes at 70 °C. The gel is then visualised by staining with SimplyBlue SafeStain (Invitrogen, Thermo Fisher Scientific).

Protein	Dialysis buffer	Final concentration
ATAD2	10 mM HEPES, 500 mM NaCl, pH 7.0, 0.5 mM TCEP, 5 % glycerol	171 μM
[^{15}N]-ATAD2	10 mM HEPES, 500 mM NaCl, pH 7.0, 0.5 mM TCEP, 5 % glycerol	65 μM
UbE2L3	20 mM HEPES, 100 mM NaCl, pH 7.5, 1 mM DTT	208 μM

Table 7.4: Final buffer conditions and concentration of purified proteins.

7.2.7. Fluorescent labelling of ATAD2

A frozen aliquot of ATAD2 in storage buffer was thawed on ice. Precipitation however, was observed after thawing, and therefore the solution was centrifuged (13000 rpm) to remove aggregated protein. The concentration was re-measured by UV/Vis spectroscopy (NanoDrop 1000 spectrophotometer (PEQLAB biotechnologie GmBH)). The Cy5-maleimide dye (50 mM stock in DMSO, final protein:dye ratio of 1:7) was added to 750 μg of ATAD2 protein in buffer (325 μl) with 0.5 mM of TCEP and left overnight at 4 °C with shaking. The next day a NAP-5

column (GE Healthcare illustra™) was equilibrated with protein buffer (10 mM HEPES, pH 7.0, 500 mM NaCl, 0.5 mM TCEP, 5 % glycerol) and the labelling mixture applied directly to the column. Fractions were collected (~200 µl) by gravity flow by applying buffer directly to the centre of the column. Separation on the column of labelled protein and free dye was observed by eye and 8 fractions were collected in total. Fractions were analysed by HPLC, method 2 (section X), and fractions 1 – 4 were pooled. To determine the concentration of labelled protein, the absorbance was measured at 280 nm and 330 nm (to correct for stray light, since folded proteins do not absorb at this wavelength, any absorbance observed at 330 nm therefore arises due to scattering) and 650 nm (optimal absorption of Cy5 dye). Sample buffer was used as a blank reference solution. The average absorbance measurements, stray light corrections, Cy5 correction factors (CF_{280} = 0.0415; CF_{330} = 0.0509) and extinction coefficients for ATAD2 (8940 $M^{-1}.cm^{-1}$) and Cy5 dye (250,000 $M^{-1}.cm^{-1}$) were used to calculate the degree of labelling according to the following equations:

$$A_{280\ corrected} = A_{280} - (A_{650} \times CF_{280})$$

$$A_{330\ corrected} = A_{330} - (A_{650} \times CF_{330})$$

$$A_{280\ stray\ light\ corrected} = A_{280\ corrected} - (A_{330\ corrected} \times 1.93)$$

Where $A_{280\ corrected}$ is the corrected absorbance at 280 nm of the protein, $A_{330\ corrected}$ is the corrected absorbance for the protein at 330 nm and $A_{280\ stray\ light\ corrected}$ is the absorbance. The total concentration of protein (M) in the sample is then given by:

$$[P]_0 = \frac{A_{280\ stray\ light\ corrected}}{8940 \times 0.1}$$

The concentration of Cy5 dye in the sample is given by:

$$[Cy5] = \frac{A_{650}}{250000 \times 0.1}$$

The ratio of the protein concentration and the dye concentration is then used to calculate the labelling efficiency. The correction factors for the free Cy5 dye have been previously calculated in the Auer lab by Dr Nhan Pham.

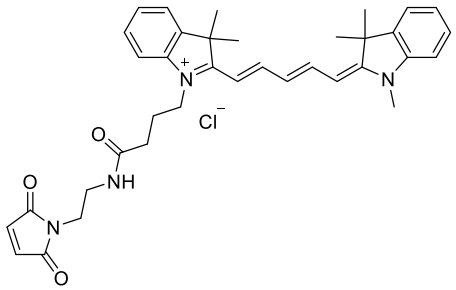


Figure 7.1: Chemical structure of Cyanine-5 maleimide dye. This dye was chosen for its favourable fluorescent properties namely, high molar extinction coefficient (250,000 $M^{-1}.cm^{-1}$).

7.2.8. Synthesis of N_α-Fmoc-N_ε-acetyl-*L*-lysine-OH

N_α-Fmoc-N_ε-acetyl-*L*-Lysine-OH was synthesised in house for use in SPPS of the H4K5 and H4 peptides.

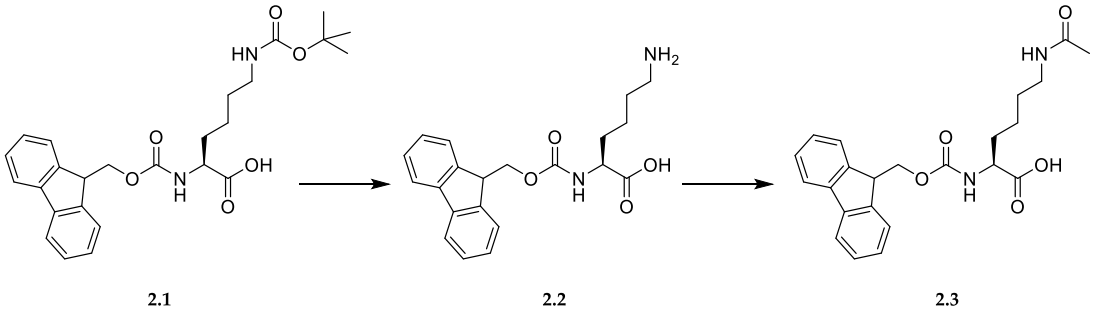
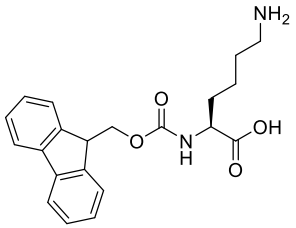


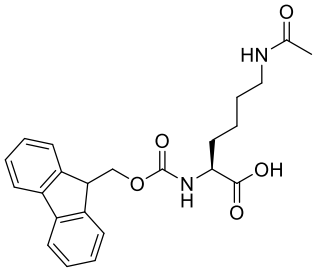
Figure 7.2: Synthetic scheme to N_α-Fmoc-N_ε-acetyl-*L*-Lysine-OH . Protocol adapted from Altamore *et al*.[253]

Synthesis of N_α-Fmoc-*L*-lysine-OH (2.2)



TFA (6.0 ml, 78.4 mmol) was added to a solution of commercially available Fmoc-*L*-Lysine(Boc)-OH (**2.1**, 2.1 g, 4.5 mmol) in DCM (60 ml). The resulting solution was stirred at room temperature for 3 hours. Thereafter, the reaction mixture was concentrated *in vacuo* on a Rotavap at 40 °C, and the residual TFA was removed by suspending the solution in diethyl ether (20 ml) followed by concentration *in vacuo* (repeated three times). The resulting residue was suspended in diethyl ether and the precipitate filtered to give title compound (**2.2**, 2.3 g, 4.8 mmol, 100 % yield) as a white solid (TFA salt). MS (ESI) exact mass calculated for **2.2** ($C_{21}H_{24}N_2O_4$) $[M+H]^+_{calc}$ 369.2; $[M+H]^+_{obs}$ 369.4.

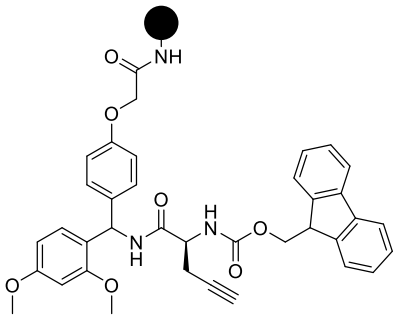
Synthesis of N_{α}-Fmoc-N_{ε}-acetyl-*L*-lysine-OH (2.3)



Ac_2O (900 μl, 9.5 mmol, 2 eq) was added to a solution of **2.2** (2.3 g, 4.8 mmol, 1 eq.) and NEt_3 (731 μl, 5.2 mmol, 1.1 eq.) in DCM (25 ml) at 0 °C and the resulting solution stirred for 2 hours. The mixture was then concentrated *in vacuo* and the residue redissolved in DCM (20 ml). The organic phase was washed with brine (10 ml), dried over Na_2SO_4, filtered and concentrated *in vacuo.* The residue was then purified by flash chromatography (Biotage ZIP® KP-Sil column (120 g), DCM:MeOH:AcOH, 96:2:2). The fractions were leuted with a gradient of 0 – 2 % methanol. The peak fractions (as analysed by TLC) were combined and concentrated under reduced pressure using a Rotavap. The residue was azeotroped three times with cyclohexane to remove excess acetic acid. The residue was then dissolved in methanol (10 ml) and purified by precipitating in an excess of water. The precipitate was filtered and dried *in vacuo* (using vacuum oven) to yield the title compound as a white solid (**2.3**, 1.7 g, 4.1 mmol, 87 %). MS (ESI) exact mass calculated for $C_{23}H_{26}N_2O_5$: $[M+H]^+_{calc}$ 411.2; $[M+H]^+_{obs}$ 411.02. HPLC (Method 1) t_R 15.9 mins; > 95 % purity as measured by peak area ^{1}H NMR (500 MHz, MeOD) δ 7.80 (2H, d), 7.69 (2H, t), 7.40 (2H, t), 7.32 (2H, t), 4.37 (2H, t) , 4.24 (1H, t), 4.16 (1H, m), 3.22 (1H, dd), 3.18 (1H, t), 1.93 (3H, s), 1.86 (1H, m), 1.72 (1H, m), 1.49 (4H, m), 1.32 (1H, t). ^{13}C NMR (126 MHz, MeOD) δ 174.5, 171.8, 157.3, 144.0, 143.8, 141.2, 127.4, 126.8, 124.9, 119.5, 66.5, 53.9, 38.8, 31.0, 28.4, 22.9, 21.2.

7.2.9. Solid-phase peptide synthesis

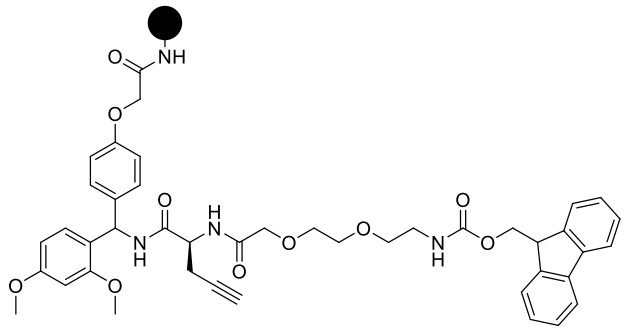
7.2.9.1. Preparation of Pra functionalised resin (TG-Pra)



TG-RAM resin was prepared by swelling 1 equivalent of dry resin (0.3 $mmol.g^{-1}$, 90 μm,) in DMF for 30 minutes with shaking. The resin was then treated twice with a solution of 20 % (v/v) piperidine in DMF at room temperature with shaking for 10 and 15 minutes respectively, with washing in between (10 column volumes of DMF). The resin was then washed thoroughly with 10 successive column volumes of DMF, DCM, MeOH and H_2O, to ensure complete removal of excess piperidine. The resin was then treated with a solution of Fmoc-Pra (6 eq.), HATU (5.5 eq.), DIPEA (12 eq.) in DMF for 30 minutes with shaking. The reaction was drained and rinsed with 10 column volumes of

DMF and the coupling reaction repeated. The Pra-functionalised resin was then washed thoroughly with DMF, H_2O, MeOH and DCM and a small amount of resin used for the TNBS test (see section 7.1.5). The resin is then dried overnight *in vacuo*.

7.2.9.2. Preparation of Pra-DOA-functionalised resin (TG-Pra-DOA)



TG-Pra resin (section 7.2.9.1) was prepared by swelling 1 equivalent of dry resin (0.3 mmol.g^{-1}, 90 μm,) in DMF for 30 minutes with shaking. The resin was then treated twice with a solution of 20 % (v/v) piperidine in DMF at room temperature with shaking for 10 and 15 minutes respectively, with washing in between (10 column volumes of DMF). The resin was then washed thoroughly with 10 successive column volumes of DMF, DCM, MeOH and H_2O, to ensure complete removal of excess piperidine. The resin was then treated with a solution of Fmoc-DOA (6 eq.), HATU (5.5 eq.), DIPEA (12 eq.) in DMF for 30 minutes with shaking. The reaction was drained and rinsed with 10 column volumes of DMF and the coupling reaction repeated. The Pra-functionalised resin was then washed thoroughly with DMF, H_2O, MeOH and DCM and a small amount of resin used for the TNBS test (see section 7.1.5). The resin is then dried overnight *in vacuo*.

7.2.9.3. Peptide H4K5 (H_2N-RGK_{ac}GG-DOA-βCF_3F-PRA-$CONH_2$)

The title peptide was synthesised by standard solid-phase synthesis methods, outlined in section 7.1.5, using prepared TG-Pra resin (100 mg, 0.3 mmol.g^{-1}). Fmoc protected amino acids (Gly, Lys(Ac), Arg(Pbf), β(CF_3)Phe) were used at 6 eq, HATU (5.5 eq.), DIPEA (12 eq.). Cleavage from the resin was carried out according to section 7.1.5. MS (ESI) exact mass calculated for $C_{42}H_{64}F_3N_{13}O_{11}$: $[M+H]^+_{calc}$ 985.4; $[M+H]^+_{obs}$ 985.6. HPLC (Method 1) gives a t_R of 15.6 mins. Absorbance is measured at 210 nm, and purity is > 90% (as measured by peak area). The final peptide was stored as a white powder at 4 °C (yield 11 mg, ~ 37 %). HPLC chromatogram and mass spectrum are given in appendix 8.1.5.

7.2.9.4. Peptide H4 (H_2N-RGK_{ac}GGK_{ac}GLGK_{ac}G-DOA-PRA-$CONH_2$)

The title peptide was synthesised by standard solid-phase synthesis methods, outlined in section 7.1.5, using prepared TG-Pra-DOA resin (200 mg, 0.26 mmol.g^{-1}). Fmoc protected amino acids (Gly, Lys(Ac), Arg(Pbf), β(CF_3)Phe) were used at 6 eq, HATU (5.5 eq.), DIPEA (12 eq.). Cleavage from the resin and purification of the peptide was carried out according to section 7.1.5 .MS (ESI) exact mass calculated for $C_{59}H_{102}N_{20}O_{18}$: $[M+H]^+_{calc}$ 1378.8; $[M+H]^+_{obs}$ 1380.8. HPLC (Method 1) gives a t_R of 11.7 mins. Absorbance is measured at 210 nm, and purity is > 90% (as measured by peak area). The peptide was stored as a white powder at 4 °C (yield 9 mg, ~ 11% yield). HPLC chromatogram and mass spectrum are given in appendix 8.1.6 .

7.2.9.5. Peptide H_2N-hoA-(*m*CF_3)F-E-R-$CONH_2$

The title peptide was synthesised by standard solid-phase synthesis methods, outlined in section 7.1.5 using TG-RAM resin (120 mg, 0.26 mmol.g^{-1}). Fmoc protected amino acids (homo-Ala, *meta*(CF_3)Phe, Arg(Pbf), Glu(OtBu)) were used at 3 eq, HATU (2.5 eq.), DIPEA (6 eq.). Cleavage from the resin and purification of the peptide was carried out according to section 7.1.5. MS (ESI) exact mass calculated for $C_{25}H_{37}F_3N_8O_6$: $[M+H]^+_{calc}$ 603.3; $[M+H]^+_{obs}$ 603.4. HPLC (Method 1) gives a t_R of 12.1 mins. Absorbance is measured at 210 nm, and purity is > 90% (as measured by peak area). The peptide was stored as a white powder at 4 °C (final yield 10 mg ~ 53 %). HPLC chromatogram and mass spectrum are given in appendix 8.1.7.

7.2.9.6. Dipeptide and tripeptide truncates: H_2N-hoA-(mCF_3)-$CONH_2$ and H_2N-hoA-(mCF_3)-E-$CONH_2$

The title peptides were synthesised by standard solid-phase synthesis methods, outlined in section 7.1.5 using TG-RAM resin (50 mg, 0.26 mmol.g^{-1}). Fmoc protected amino acids (homo-Ala, *meta*(CF_3)Phe, Arg(Pbf), Glu(OtBu)) were used at 3 eq, HATU (2.5 eq.), DIPEA (6 eq.). Cleavage from the resin and purification of the peptide was carried out according to section 7.1.5. MS (ESI) exact mass calculated for: H_2N-hoA-(mCF_3)-$CONH_2$ $[M+H]^+_{calc}$ 603.3; $[M+H]^+_{obs}$ 603.4. HPLC (Method 1) gives a t_R of 12.1 mins. Absorbance is measured at 210 nm, and purity is > 90% (as measured by peak area). The peptide was stored as a white powder at 4 °C (final yield 10 mg ~ 53 %). HPLC chromatogram and mass spectrum are given in appendix 8.1.7.

7.2.9.7. Fluorinated Dipeptide and Tripeptide Library synthesis

The fluorinated dipeptides, and C- and N-terminal tripeptides were synthesised by standard solid-phase synthesis methods, outlined in section 7.1.5. Each peptide was synthesised using

TG-RAM resin (25 mg; 0.23 mmol.g^{-1}). After cleavage and drying overnight, peptides were dissolved in 500 µl 25 % ACN in H_2O and analysed by HPLC and LCMS. Peptides were then flash frozen in liquid N_2 and dried overnight in a freeze drier (Christ, Alpha 1-2 LD Plus) operating at 0.2 mbar and – 52 °C (for the ice condenser). Following this, solid peptides were weighed and DMSO added to a final concentration of 10 mM. Peptides are stored at -20 °C as DMSO stock solutions and 4 °C as a white powder. The chemical structures, HPLC and LCMS QC data for all fluorinated di- and tripeptides are given in appendix 8.1.10-12.

7.2.9.8. Solid-phase Synthesis of OBOC Lysine-PTM Library

The SCAL2-PRA-DOA functionalised resin was synthesised using standard solid-phase synthesis methods. SCAL2 was prepared in-house by Peter Dodd and used as needed.

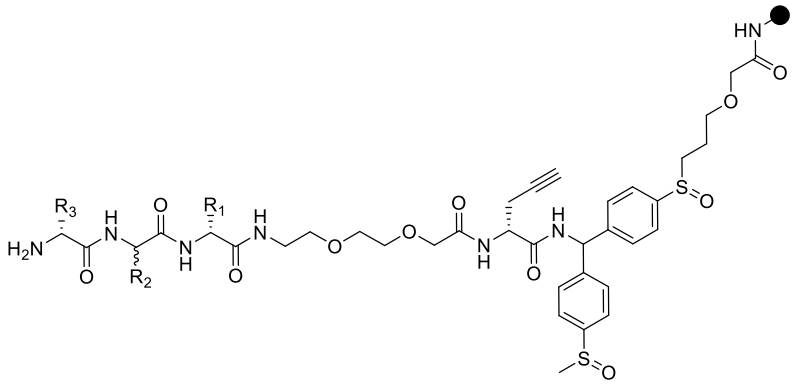


Figure 7.3: Chemical structure of OBOC library. SCAL-2 functionalised resin and R1 = randomised *L* amino acid; R2 = randomized lysine derivative amino acid; R3 = fixed position *L* amino acid/sub-library).

Using the functionalized resin, tripeptide sub-libraries were prepared by the solid-phase method on a 365 mg (110 µmol) scale following the Fmoc strategy and using standard Fmoc-derivatized amino acids (6 eq). Activation of amino acids was achieved using HATU-DIPEA (5.5 and 12 eq. respectively), and Fmoc deprotection was carried out using a 20% (v/v) piperidine solution in DMF. All couplings were performed for 2 x 30 minutes and deprotections for 1x10 + 1x15 minutes. Fixed positions (R3) were introduced using 6 eq of a single protected amino acid derivative. Building blocks for randomized positions (R1, R2) were coupled using 4 or 6 eq of a single protected amino acid derivative (see table 7.5 and 7.6 respectively) then a split and mix strategy was applied. The resin was split into 16 SPEs (~20 mg per sublibrary, 6.6 µmol) and the first amino acid coupling performed in parallel. Following thorough washing steps beads were thoroughly mixed in a SPE (10 ml volume) by inverting several times. The bead mixture was deprotected and then split further into 12 sub-

libraries. The second amino acid coupling was performed, mixed and deprotected as before. The resin was mixed thoroughly again and then split into the final 16 sub-libraries. The final amino acid coupling was performed and the resin washed and dried *in vacuo*, then stored under reduced pressure in a desiccator prior to QC analysis.

Syringe #	Name	MW	Density	eq	mmol	mg per coupling	µl per coupling
	resin(NH_2)	0.3		1	0.006	20	
	HATU	380.2		5.5	0.033	12.5	
	DIPEA	129.2	0.742	12	0.072		12.6
	DMF	73.1	0.948				
1	*L*-ALA	311.3		6	0.036	11.2	
2	*L*-ARG(pbf)	648.8		6	0.036	23.4	
3	*L*-ASN(trt)	596.7		6	0.036	21.5	
4	*L*-ASP(Otbu)	411.5		6	0.036	14.8	
5	*L*-GLN(trt)	610.7		6	0.036	22.0	
6	*L*-GLU(Otbu)	425.5		6	0.036	15.3	
7	*L*-HIS(trt)	619.7		6	0.036	22.3	
8	*L*-LEU	353.4		6	0.036	12.7	
9	*L*-LYS(boc)	468.5		6	0.036	16.9	
10	*L*-PHE	387.4		6	0.036	13.9	
11	*L*-PRO	337.4		6	0.036	12.1	
12	*L*-SER(Otbu)	383.4		6	0.036	13.8	
13	*L*-THR(Otbu)	397.5		6	0.036	14.3	
14	*L*-TRP(boc)	526.6		6	0.036	19.0	
15	*L*-TYR(Otbu)	459.6		6	0.036	16.5	
16	*L*-VAL	339.4		6	0.036	12.2	

Table 7.5: Equivalents and amounts of *L*-amino acids used for the coupling reaction in construction of the OBOC library (coupling reactions were performed twice). This was used for coupling reactions at the R1 and R3 positions in the final library (figure 7.3).

Syringe #	Name	MW	Density	eq	mmol	mg per coupling	µl per coupling
	resin(NH_2)	0.3		1	0.006		
	HATU	380.2		5.5	0.033	12.5	
	HATU	380.2		3.5	0.021	8.0	
	DIPEA	129.2	0.742	12	0.072		12.6
	DIPEA	129.2	0.742	8	0.048		8.4
	DMF	73.1	0.948				
1	Lys-DEV-1	410.5		6	0.036	14.8	
2	Lys-DEV-2	464.4		6	0.036	16.7	
3	Lys-DEV-3	472.5		6	0.036	17.0	
4	Lys-DEV-4	584.6		4	0.024	14.0	
5	Lys-DEV-5	480.6		4	0.024	11.5	
6	Lys-DEV-6	480.6		4	0.024	11.5	
7	Lys-DEV-8	482.6		4	0.024	11.6	
8	Lys-DEV-10	587.7		4	0.024	14.1	
9	Lys-DEV-11	601.7		4	0.024	14.4	
10	Lys-DEV-12	473.5		6	0.036	17.0	
11	Lys-DEV-15	482.6		4	0.024	11.6	
12	Lys-DEV-16	524.6		6	0.036	18.9	

Table 7.6: Equivalents and amounts of Lysine-DEV (DEV = derivative) amino acids used in the coupling reaction at position R2 of the OBOC library (figure 7.3). Lysine-DEV structures are given in appendix 8.1.9. Lys-DEVs 7, 9, 13, 14 failed QC checks, therefore were not included in the final library.

Library Summary:

- Solid support: SCAL linker functionalised TentaGel™ beads (0.3 $mmol.g^{-1}$loading capacity, 140 µm). 20 mg resin per sub (theoretical 6.6 µmol).
- 3072 member compound library, 16 x 12 x 16 tripeptide sequences in 16 sub-libraries. After the last amino acid coupling, the resins are retained in 16 SPE vials, this means that the last residue of the library sequence is known, hence 16 sub-libraries. Therefore, screening pools of 192 compounds per sub-library.
- 16 x FMOC protected *L*-amino acids: *L*-ALA, *L*-ARG(pbf), *L*-ASN(trt), *L*-ASP(Otbu), *L*-GLN(trt), *L*-GLU(Otbu), *L*-HIS(trt), *L*-LEU, *L*-LYS(boc), *L*-PHE, *L*-PRO, *L*-SER(Otbu), *L*-THR(Otbu), *L*-TRP(boc), *L*-TYR(Otbu), *L*-Val.
    - 12 x FMOC protected lysine derivative amino acids: Nα-Fmoc-Nε-acetyl-L-Lysine, Nα-Fmoc-Nε-trifluoroacetyl-L-Lysine, Nα-Fmoc-Nε-nicotinyl-L-Lysine, Nα-Fmoc-Nε-7-methoxycoumarin-4-acetyl-L-lysine, Fmoc-(1-Boc-piperidin-3-yl)-DL-glycine, Nα-Fmoc-b-(1-Boc-piperidin-4-yl)-DL-alanine, Nα-Fmoc-Nε-methyl-Nε-Boc-L-

Lysine, Nα-Fmoc-Nε-(Boc-2-aminobenzoyl)-L-Lysine, Nα-Fmoc-Nε-dansyl-L-Lysine, Nα-Fmoc-Nε-nicotinyl-D-Lysine, Fmoc-N-methyl-Lysine(Boc), (2S,5R)-Fmoc-2-amino-4-(3-Boc-2,2-dimethyl-oxazolidin-5-yl)-butyric acid (structures given in appendix 8.1.9).

7.2.9.9. QC analysis of library

For purity and loading determination of the OBOC library a small amount of resin (~1 mg) from each sub-library was placed into a glass vial and labelled *via* the N-terminus with TMR-acid (6-TAMRA). The 6-TAMRA is used exactly as an amino acid and coupling reaction performed as described in section 7.1.5. The coupling reaction was performed in a glass MS vial for ease of bead picking. After the labelling reaction, the resin was thoroughly washed and dried overnight *in vacuo*. The 16 vials containing the TMR labelled beads were treated with an excess of TFA (200 μl), twice for 1 hour each with shaking. The TFA was then removed under reduced pressure in a Speed Vac and the 16 batches were suspended in 50 % ACN in H_2O (1 ml) and left overnight at 4 °C. Following this, the supernatant was removed and single beads were picked using a Gilson pipette, with 10 μl pipette tip and placed into a separate vial. The vial containing a single bead were treated with a 12 %(v/v) solution of TMSBr in TFA for 2 hours. The TFA was removed under reduced pressure in a speed vac and 20 μl of a 50 % ACN solution in H_2O was added. The solution is left to stand for 15 minutes prior to injection for HPLC analysis (injection volume, 15 μl). HPLC analysis was performed using Method 1 (section 7.1.2) and fluorescence detection at λ_{ex} 555 nm, and λ_{em} 575 nm. The purity was calculated from the percentage peak area of the labelled peptide and the loading capacity is calculated from the TAMRA standard curve (shown in appendix 8.1.8).

7.2.10. [^{1}H, ^{15}N]-HSQC Experimental analysis of ATAD2 with peptide H4

The synthesised peptide H4 was dissolved in NMR buffer (10 mM HEPES, 500 mM NaCl, 2 mM DTT, pH 7.0, 10 % D_2O) to a final stock concentration of 10 mM. [^{15}N]-ATAD2 (65 μM) was analysed by 2D-HSQC NMR experiments on the AVA 800 MHz spectrometer. The starting volume was 400 μl in a 5 mm NMR tube (Norell® standard series). Spectra were acquired with the following parameters: spectral width 27 ppm, time domain 256 (in F1), number of scans 8, receiver gain 128 and D1 = 1.4 s. A titration with peptide H4 according to table 7.7 was performed and spectra analysed with NMRViewJ.

Titration Point	$[P]_0$ final (μM)	$[L]_0$ final (μM)	Molar Ratio (L:P)
1	65	0	0
2	59.7	50	0.83
3	59.4	100	1.68
4	58.8	200	3.40
5	57.7	400	6.93
6	54.5	1000	18.35

Table 7.7: Experimental set-up of [^{15}N]-HSQC ATAD2 titration with peptide H4.

7.2.11. OPPG Proof of concept experiments with H_2N-hoA-(*m*CF_3)F-E-R-$CONH_2$

The survivin protein was supplied by Dr Nhan Pham (Auer group) at a concentration of 134 μM in buffer (20 mM HEPES, 100 mM NaCl, pH 7.5). A stock solution of peptide was made in non-deuterated DMSO (final concentration 50 mM). Two samples were prepared in NMR buffer (20 mM HEPES, 100 mM NaCl, pH 7.5, 10 % D_2O) with final concentration of peptide at 100 μM (final DMSO concentration 0.2 %). To one sample, survivin was added to a final concentration of 10 μM and to the other, equal volume of NMR buffer. The ^{19}F NMR spectra were recorded (see section 7.1.3.1) and analysed in MestReNova (see section 7.1.4).

7.2.12. Fluorine Dipeptide/Tripeptide Screening against UbE2L3

Fluorinated dipeptides were screened against UbE2L3 using the ^{19}F NMR experimental parameters described in section 7.1.4. Peptides were pooled based on their minimal ^{19}F resonance overlap (peptide pools are given in appendix 8.1.10). The peptides were dissolved in NMR assay buffer (20 mM HEPES, 100 mM NaCl, pH 7.5, 10 % D_2O) to a final concentration of 100 μM of each peptide (final DMSO concentration of pool is 4 %) and final volume of 1.1 ml. The peptide pools were aliquoted into two separate Eppendorf tubes then protein was added to one at a final concentration of 10 μM and an equivalent amount of buffer added to the control. Samples were pipetted into NMR tubes and placed in the sample changer of the AVA 400 MHz spectrometer for analysis. 1D ^{19}F NMR spectra were processed using MestReNova and the chemical shifts recorded in a table. The ^{19}F chemical shifts (in Hz) from the protein containing sample were subtracted from the control sample and the data plotted as a bar chart in Microsoft Excel (these are illustrated in chapter 2).

7.3. Experimental Methods for Chapter 4

7.3.1. Fragment compound handling and storage

Fragment compounds were purchased from various commercial vendors, specifically: Fluorochem Ltd, Enamine Ltd, Vitas M Chemical Ltd and Selleck Chemicals.

Stock solutions of each commercial fragment compound were constructed by weighing 1 - 5 mg of compound and adding the calculated amount of non-deuterated, anhydrous DMSO to a final fragment concentration of either 10 or 100 mM, where stated. Solubility was confirmed by visual inspection. For compounds that were liquids, the compound density was used to calculate the amount of DMSO required for a final concentration 100 mM using a 5 µl aliquot of compound. DMSO fragment stocks were stored at - 20 °C and solid fragments stored at 4 °C. Quality control of the fragments was performed using three orthogonal standard methods: HPLC, MS (ESI) and ^{19}F-NMR.

Sample preparation for HPLC and LCMS: Fragment stocks were diluted into 25 % ACN to a final concentration of 100 µM (1:1000 dilution of 100 mM stock). Samples were analysed using the HPLC method 1 and LCMS methods described (section 7.1.2) with an injection volume of 25 µl and 4 µl, respectively.

Sample preparation for ^{19}F NMR: Soluble fragments, from DMSO stocks, were further diluted to 100 µM in NMR minimal buffer (20 mM HEPES, 100 mM NaCl, pH 7.5, 10 % D_2O) and their individual 1D ^{19}F NMR spectrum recorded using the AVA 400 MHz spectrometer as described in section 7.1.3.1.

The chemical structures and QC data for all fragments are given in appendix 8.3.3-5.

7.3.2. HSA Experimental Methods

HSA (Human Serum Albumin, recombinant, expressed in rice, A9731, Sigma Aldrich), was supplied as a lyophilised powder and dissolved in NMR assay buffer (20 mM HEPES, 100 mM NaCl, pH 7.5, 10 % D_2O) to a final concentration of 300 µM (monomer M_w ~67 kDa). Fragment stock solutions of 3-fluoro-2-hydroxy benzoic acid (FHBA), 4-(trifluoromethyl)benzylamine (TFBA), 4-(fluoro)benzylamine (FBA) were made by dissolving the compounds in 100 % DMSO at a final concentration of 100 mM. A 1D ^{19}F NMR spectrum of FHBA was recorded according to the parameters set out in section 7.1.3.1, with a varying number of scans, ranging

from: 128, 256, 512, 1024 and 2048. Spectra were processed and analysed using MestReNova NMR software. For the HSA concentration series of experiments, 6 samples were prepared in NMR buffer (100 μM FHBA, 20 mM HEPES, 100 mM NaCl, pH 7.5, 10 % D_2O), then HSA was added to a final concentration of: 0 M, 100 nM, 500 nM, 1 μM, 2.5 μM and 5 μM. The ^{19}F NMR spectra were recorded as described in section 7.1.3.1 and spectra processed in the same way.

7.3.3. UbE2L3 enzymatic activity assay (Performed with Dr. Joanna Koszela)

The reagents for the gel-based ubiquitination assay (see table 7.8) were prepared on ice. A master mix for volume of 5 wells was made containing 5 μM ubiquitin and 50 nM E1 enzyme in ubiquitination buffer (50 mM Tris, pH 7.5, 5 mM $MgCl_2$).

Reagent	Stock Concentration	Final Concentration	Per well (μl)	For 5 X wells (μl)
H_2O			16.58	82.89
10x buffer	10	1	2.20	11.00
WT Ub	100	5	1.10	5.50
E1	5	0.05	0.22	1.10
UbE2L3	137	5	0.80	
ATP (mM)	100	5	1.10	
		Total Volume	**22**	**100.5**

Table 7.8: Reagents for the SDS-PAGE assay were prepared on ice. A master mix with enough material for 5 experiments was made to maintain equal concentrations of E1 and ubiquitin across samples.

The master mix was split between 4 microtubes in 20.1 μl aliquots and 0.8 μl of UbE2L3 (137 μM) or buffer was added according to the appropriate tubes. Also, 1.1 μl of ATP (100 mM in H_2O) was added to the appropriate tubes for a final concentration of 5 mM. The reactions were then incubated at 37 °C for 2 hours and the reactions quenched by addition of 8 μl SDS loading buffer (Bolt® LDS sample loading buffer) and 2 μl H_2O or 2 μl DTT (1 M) (table 7.9). The samples were resolved by SDS-PAGE and visualised by Coomassie staining (SimplyBlue™, Invitrogen). The results are presented in the main text (chapter 4.2.5.1).

Lane	2	3	4	5
Buffer	+	+	+	+
Ub	+	+	+	+
E1	+	+	+	+
E2	+	-	+	+
ATP	-	+	+	+
DTT	-	-	-	+

Table 7.9: SDS-PAGE gel lane map of components in ubiquitination assay.

7.3.4. Compound pools and sample preparation

A ^{19}F spectrum of each fragment in aqueous buffer (20 mM HEPES, 100 mM NaCl, pH 7.5, 10 % D_2O) at 100 μM was measured and the chemical shift recorded in a table. Fragment library pools were constructed by minimal overlap analysis of ^{19}F-NMR chemical shifts. The table was then sorted by chemical shift from lowest to highest. The lowest compound was taken and placed into pool 1, the second compound placed into pool 2 etc until the desired number of pools had been reached. Fragment pools are given in appendix 8.3.3 - 5.

For screening purposes, the volume of the solution has to be 500 μl (the AVA 400 spectrometer has a 5 mm probe and using less volume will make the samples difficult to shim and lead to broader peaks). Care must be taken to maintain equal concentrations of fragments between protein and non-protein containing pools, as minute changes in DMSO concentration and pH can affect the resonance signal. A 1.1 ml fragment pool solution is made up (for example, 1.1 μl of 100 mM fragment stock is added to 1.1 ml of NMR buffer, to give final fragment concentration of 100 μM). The fragment pool is then divided equally into two Eppendorf tubes and the required amount of protein volume is added to the protein-containing pool and protein dialysis buffer added to the other Eppendorf tube. This maintains equal amounts of fragment and DMSO concentration, between the control and protein-containing samples therefore any changes in the spectra can be interpreted as a ligand binding interaction. The samples in Eppendorf tubes are centrifuged briefly, then pipetted directly into 5 mm NMR tubes. The tubes are placed in the sample rack of the AVA 400 NMR spectrometer and the 1D ^{19}F NMR spectra recorded (see section 7.1.3.1).

7.3.5. Fragment screening of UbE2L3

Fragment pools were prepared as described (section 7.3.4) in UbE2L3 buffer (20 mM HEPES, 100 mM NaCl, pH 7.5, 10 % D_2O). The fragments pools were analysed by ^{19}F NMR (section 7.1.3.1) in the presence and absence of 10 μM UbE2L3. NMR spectra were analysed by overlaying the pools with and without protein for indications of chemical shift changes and/or line broadening using MestReNova. The chemical shift changes were calculated by subtracting the protein containing ^{19}F resonance (in Hz) from the control sample. The data was then plotted as a bar chart using Microsoft Excel.

7.4. Experimental Methods for Chapter 5

7.4.1. NusEΔ/NusB Protein Handling

The NusEΔ/NusB protein was supplied as a lyophilised powder in buffer (25 mM Tris.HCl, 150 mM NaCl, pH 7.5). A 500 μl aliquot of deionised H_2O was added to the Eppendorf tube and the sample left to stand for 2 minutes. The solution was mixed by gentle pipetting and then centrifuged for 5 minutes at 4 °C and 13200 rpm in a table-top centrifuge. The supernatant was transferred to a new Eppendorf and the concentration measured by UV-Vis spectroscopy (NanoDrop 1000 spectrophotometer (PEQLAB biotechnologie GmbH)). The details and parameters for concentration determination of the NusE/NusB complex are given in table 7.10. The protein was divided into 30 μl aliquots at 208 μM and flash frozen in liquid N_2 for storage at – 80 °C.

The Δ represents a deletion of a 21 amino acid loop (and replaced with a single serine) that has been shown to increase the solubility and stability of the NusE/NusB complex, however, the K_D of the NusEΔ/NusB complex in this form is not known. The deleted region is thought to be an RNAP binding region.

	NusEΔ
Sequence	GPLGSMQNQRIRIRLKAFDHRLIDQATAEIVETAKRTGA QVRGPIPLPTRSRTHLRLVDIVEPTEKTVDALMRLDLAAG VDVQISLG (residues 46-67 replaced by a Ser, highlighted in red).
No. of amino acids	87
Molecular Weight (Da)	9593.1
Theoretical pI	10.2
Extinction coefficient (M^{-1}. cm^{-1})	N/A
	NusB
Sequence	MGSSHHHHHHSSGLVPRGSHMKPAARRRARECAVQAL YSWQLSQNDIADVEYQFLAEQDVKDVDVLYFRELLAGV ATNTAYLDGLMKPYLSRLLEELGQVEKAVLRIALYELSKR SDVPYKVAINEAIELAKSFGAEDSHKFVNGVLDKAAPVI RPNKK
No. of amino acids	159
Molecular Weight (Da)	17852.4
Theoretical pI	7.9
Extinction coefficient (M^{-1}. cm^{-1})	15930

Table 7.10: NusEΔ/NusB parameters as calculated by ExPASy ProtParam. The molar extinction coefficient here was used to determine protein concentration of the complex.

7.4.2. ^{19}F NMR Fragment Screening

Fragment pools were prepared as described (section 7.3.4) in NusEΔ/NusB buffer (25 mM Tris/HCl, 150 mM NaCl, pH 7.5, 10 % D_2O). The fragments pools (100 μM each compound) were analysed by ^{19}F NMR in the presence and absence of 10 μM protein as described in section 7.1.3.1.

7.4.3. Experiments to Validate Fragment Binding

NMR samples were prepared at 100 μM and 200 μM fragment concentration (500 μl volume) in NMR buffer (25 mM Tris/HCl, 150 mM NaCl, pH 7.5, 10 % D_2O) for the controls, and another sample was prepared including 5 μM protein. The dCSP measurements were conventional 1D ^{19}F NMR spectra (recorded according to section 7.1.3.1). The chemical shifts were analysed by MestReNova software and K_D calculated using equation 4.10.

The CPMG T_2 filtered experiments were a modified version of the 1D ^{19}F NMR experiment and utilised a spin echo period of the form [D1-90_x-(Δ-180_y- Δ)$_n$-acquire] before acquisition

where D1 = 8.5 s, Δ = 10 ms, n = 2, 4, 8, 20, 40 (for a total SL time of $2n \times 10$, see table 7.11). The data were collected with a SW of 22727 Hz and 256 scans. The absolute intensities of the fluorine resonance were analysed using MestReNova and plotted against the total length of the spin lock period in ms (data were plotted using Origin 9.0).

n	CPMG Spin Lock (ms)
2	40
4	80
8	160
20	400
40	800

Table 7.11: Number of 180 ° pulses (n) in the spin echo period of the CPMG pulse sequence. The spin lock time (ms) is therefore calculated by: $2n \times \Delta$. Where Δ is the delay between pulses (10 ms).

7.4.4. Fragment NOESY experiments

2D homonuclear ^{19}F-^{19}F NOESY and ^{1}H-^{1}H NOESY spectra were recorded on the PRO 500 MHz spectrometer and AVA 800 MHz spectrometer respectively. Samples were prepared in NMR buffer (500 µl final volume; 25 mM Tris/HCl, 150 mM NaCl, pH 7.5, 10 % D_2O; final DMSO concentration 0.02 %) using 100 µM CFL02 and CFL17, and 5 µM protein concentration, also a no-protein containing control sample was also measured. Spectra were recorded with the parameters laid out in table 7.12.

Parameter	^{19}F-^{19}F NOESY (SFO1 470 MHz)	^{1}H-^{1}H NOESY (SFO1 799 MHz)
t1 increments	256	1024
NS	32	32
SW (ppm)	10	11
Mixing time (ms)	0.5	0.6

Table 7.12: NMR parameters for homonuclear NOESY experiments.

7.4.5. CFL analogue QC (HPLC, ^{1}H, ^{13}C NMR)

Compound analogues (CFL02a, CFL02b, CFL02c, CFL02d, CFL02e, CFL02f, and CFL17a) were purchased from Fluorochem Ltd (all chemical structures are shown in appendix X), and QC analysis was performed using HPLC, to determine purity, and NMR for identification. Stock compound solutions were made up at 100 mM concentration in 100 % DMSO. For HPLC analysis compounds were diluted in 25 % ACN to a final concentration of 100 μM and 0.1 % DMSO concentration (method 1). For identification by NMR, compounds were dissolved in deuterated DMSO-d_6 at a concentration of 10 $mg.ml^{-1}$ and analysed by ^{1}H and ^{13}C NMR using the AVA 500 spectrometer. QC data is given in appendix 8.4.2.

7.4.6. Bacterial Inhibition Assay

Assay optimisation

To optimise the assay, two parameters, the starting cell OD and final DMSO concentration, were adjusted for the best results. A preliminary experiment was performed with starting OD measurements at 0.01, 0.005, 0.001 and 0 (test control) AU. DMSO concentration optimisation is necessary as compound fragments are routinely stored at 100 mM in 100 % DMSO. Final DMSO concentration for optimisation studies ranged from 0, 1, 2, 3 and 4 %.

A 50 μl aliquot of *E. coli* BL21(DE3) cells was thawed and inoculated into 5 ml of LB broth. The cell suspension was then incubated overnight at 37 °C with shaking (180 rpm). A 1 ml aliquot was taken and the optical density (OD) was measured at 600 nm wavelength (Beckman DU 530 Life Science UV/Vis Spectrophotometer). Cells were diluted accordingly to final OD concentrations and aliquoted (100 μl) to a 96 well plate (Whatman, Glass Bottom, Polystyrene Microplate). DMSO was added to the final concentration as above. The plate was then inserted into a plate reader (TECAN infinite M200 PRO) and incubated at 37 °C. The plate was shaken (orbital amplitude 2 mm) for 800 s prior to measurement at 600 nm wavelength, with a settle time of 100 ms and measurements were taken every 15 minutes. The data is plotted with Microsoft Excel and the growth curves are shown in appendix 8.4.4.

Final assay conditions

A 50 μl aliquot of *E. coli* BL21(DE3) cells was thawed and inoculated into 5 ml of LB broth. The cell suspension was then incubated overnight at 37 °C with shaking (180 rpm). A 1 ml aliquot was taken and the OD was measured at 600 nm wavelength. Cells were then diluted to OD

0.01 in LB medium. Fragment hit compounds (CFL02, CFL02f, CFL17 and CFL17a) were dissolved in LB medium to final concentrations of 100 μM, 500 μM and 1 mM with a final DMSO concentration of 1 % in all samples and a blank sample as a control. Also, compound CFL23, which did not show binding activity in the ^{19}F NMR assay was used as a negative control compound. Compounds were plated in 100 μl aliquots on a 96 well plate in triplicate. A 10 μl aliquot of cells was then added to each well, to give a final OD of 0.001. Negative controls were included with no cells and compound only. The plate was measured using the same script as above.

7.4.7. Quantitative Equilibrium Microdialysis (qμD)

Quantitative equilibrium microdialysis (qμD) was used as an orthogonal, complimentary assay to confirm binding of the hit fragments and compound analogues.

Calibration curve and dialysis control

A calibration curve for compounds CFL02, CFL02f, CFL17 and CFL17a was achieved by constructing a series of increasing compound concentrations to determine the linear range of the spectrometer. Compound samples containing 10 nM, 50 nM, 100 nM, 500 nM, 1 μM, 5 μM and 10 μM were made up in assay buffer (20 mM Tris/HCl, 50 mM NaCl, 2 % DMSO and 0.1 % pluronic acid). The compounds were then tested using the LCMS method described below. Following this, compounds were tested for their ability to dialyse across the membrane. Rapid equilibrium device (RED) inserts (MWCO 8 kDa) were placed in the well of a compatible base plate (P89811, Thermo Fisher Scientific) and 100 μl of 40 μM compound in assay buffer (20 mM Tris/HCl, 50 mM NaCl, 2 % DMSO and 0.1 % pluronic acid) was added to the sample chamber (red compartment, V_1) and 300 μl of assay buffer was added to the dialysis chamber (white compartment, V_2). The plate was sealed with PlateSealing Tape for 96-well plates (15036, Thermo Fisher Scientific) and incubated at room temperature overnight (~16 hours) with rotation (300 rpm). Following overnight incubation, 50 μl of sample is removed from each compartment and added to a MS vial (VWR, 300 μl). The samples were run using LCMS analysis described below.

Equilibrium Microdialysis with CFL02f

Equilibrium microdialysis was used for determination of the dissociation constant of fragment hit CFL02f against the NusEΔ/NusB protein complex. Rapid equilibrium device (RED) inserts

(MWCO 8 kDa) were placed in the well of a compatible base plate (P89811, Thermo Fisher Scientific). NusEΔ/NusB protein was diluted to 40 μM in assay buffer (20 mM Tris/HCl, 50 mM NaCl, and 0.1 % pluronic acid) and compound was added to a final concentration of 40 μM with a final DMSO concentration of 2 %. A 100 μl aliquot of protein-ligand solution was added to the sample chamber (red compartment) and 300 μl of assay buffer added to the dialysis chamber (white compartment). The plate was sealed with PlateSealing Tape for 96-well plates (15036, Thermo Fisher Scientific) and incubated at room temperature overnight (~16 hours) with rotation (300 rpm). The dialysis experiment was performed in triplicate for the control (no protein) and compound containing samples. Following this, a 10 μl aliquot from each sample chamber was taken and diluted 1:20 into assay buffer, containing 0.1 % formic acid, in a MS vial. The samples were then analysed by LCMS on an Orbitrap system.

LCMS analysis

The vials were placed in the refrigerated (10 °C) sample injector drawer of the Acquity UPLC (Waters) and 2 μl of sample injected onto the C18 HPLC column. The solvent system for elution is: Buffer A – H_2O (0.1 % FA). Buffer B – ACN (0.1 % FA). The standard gradient program used for analysis was: 0 – 3 min, 0.5 % B; 3 – 18 min, 70 % B; 18 – 22 min, 95 % B; 22 – 25 min 0.5 % B; with a flow rate of 8 $\mu l.min^{-1}$. The mass analyser is an LTQ Orbitrap (Thermo Finnigan) with ESI injection. This method was used to quantify the peak intensity response as a function of concentration in both compartments of the RED apparatus following dialysis. Thermo Xcalibur software was used to analyse the mass spectra and calculate intensities for generation of calibration curves and K_D determination from equilibrium microdialysis. The compound CFL17a did not produce a signal, indicating that the compound did not ionise.

7.4.8. NusEΔ/NusB HSQC Titrations

^{15}N-labelled NusEΔ/NusB protein was kindly supplied by the Department of Biopolymers at the University of Bayreuth. When required, a 500 μl aliquot of approximately 200 μM protein was thawed and transferred to a 2 ml micro test tube. A piece of dialysis membrane (MWCO 3 kDa) was placed over the top and the tube placed upside down into dialysis buffer (25 mM Tris/HCl, 150 mM NaCl, pH 7.5). The protein was left to dialyse overnight at 4 °C with gentle mixing. The protein was then centrifuged briefly and the concentration measured by UV-Vis spectroscopy (extinction coefficient given in table X). Dialysed protein was stored at 4 °C and this process was repeated when necessary.

NMR samples of 550 µl in volume were prepared for 2D ^{15}N-HSQC experiments. The final protein concentration was 100 µM in NMR assay buffer (25 mM Tris/HCl, 150 mM NaCl, 10% D_2O, 0.01 % NaN_3). Compound fragments were dissolved in 100 % DMSO to final concentrations of 10 and 25 mM. Compound titrations for fragments CFL02, CFL02a, CFL02b, CFL17 and CFL17a were completed according to table 7.13. Spectra were acquired with the following parameters: spectral width 20 ppm, time domain 256 (in F1), number of scans 16, receiver gain 912 and D1 = 1s.

Titration Point	$[P]_0$ final (µM)	$[L]_0$ final (µM)	Molar Ratio ([L]:[P])	Final DMSO %	Cumulative Volume of Ligand (µl)	Final Volume (µl)
1	100	0	0	0	0	550.0
2	99.55	45.25	0.45	0.45	2.5	552.5
3	98.57	143.37	1.45	1.43	8	558
4	97.26	274.09	2.82	2.74	15.5	565.5
5	95.07	492.65	5.18	4.93	28.5	578.5
6	93.46	909.09	9.73	6.54	38.5	588.5
7	89.65	1890.79	21.09	10.35	63.5	613.5

Table 7.13: The titration points used for compounds CFL02, CFL02a, CFL02b, CFL17 and CFL17a. $[P]_0$ final protein concentration; $[L]_0$ final ligand concentration.

For each titration, the compound was added directly to the NMR tube. The tube was then inverted several times and shaken vigorously. The sample was then centrifuged in a manual, table-top centrifuge to ensure all the sample mixed thoroughly and at the bottom of the tube. The compound titration for CFL02f was carried out according to table 7.14:

Titration Point	$[P]_0$ final	$[L]_0$ final	Molar Ratio	Final DMSO %
1	100	0	0	1
2	99.55	45.25	0.45	1.45
3	99.10	90.09	0.91	1.90
4	98.65	134.53	1.36	2.35
5	98.21	178.57	1.82	2.79
6	97.78	222.22	2.27	3.22
7	96.92	308.37	3.18	4.08
8	96.07	524.017	5.45	4.93
9	95.24	735.93	7.73	5.76
10	94.42	944.21	10	6.58
11	90.53	1934.16	21.36	10.47

Table 7.14: Titration data for CFL02f. $[P]_0$ final protein concentration; $[L]_0$ final ligand concentration.

A DMSO control experiment was also carried out according to table 7.15.

Titration Point	$[P]_0$ final	Final DMSO %
1	100	0
2	99.55	0.45
3	99.10	1.43
4	98.65	1.87
5	98.21	2.74
6	97.78	4.01
7	96.92	4.93
8	96.07	6.54
9	95.24	10.35

Table 7.15: Titration data for DMSO. The final DMSO concentration used was to equal the concentration used in compound titrations.

For K_D determination and binding site mapping the backbone assignments were taken from previous studies in the Biopolymer group, University of Bayreuth.[237] To evaluate HSQC titration experiments, the calculated normalised chemical shift changes can be quantified for a specific resonance at each titration point ($\Delta\delta_{norm}$) using the formula:

$$\Delta\delta_{norm} = \sqrt{(\Delta\delta^1H)^2 + (0.1 \times \Delta\delta^{15}N)^2}$$

Equation 4.11

Where $\Delta\delta^1H$ and $\Delta\delta^{15}N$ are the chemical shift changes of amide proton and nitrogen, respectively. The '0.1' constant is a scaling factor used to balance the contributions of ^{1}H and ^{15}N to the magnitude of the perturbation.[178] As a control, the normalised chemical shifts for the DMSO titration are subtracted from the compound containing titrations. This is to correct for changes in backbone residues that are caused by changes in DMSO concentration. Dissociation constants can then be obtained by monitoring these changes over the course of a titration at different ligand concentrations. The fitting is performed using equation 4.12 (using Grafit v 7.0) for each amide resonance showing fast exchange on the chemical shift timescale.

$$\Delta\nu_{obs} = \Delta\nu_{max}\frac{([P]_0 + [L]_0 + K_D) - \sqrt{([P]_0 + [L]_0 + K_D)^2 - (4[P]_0[L]_0)}}{2[P]_0}$$

Equation 4.12

Where $\Delta\nu_{obs}$ is the normalised resonance frequency difference in Hz and $\Delta\nu_{max}$ is the normalised resonance frequency between the free and fully bound protein in Hz, K_D is the dissociation constant, $[P]_0$ the total concentration of ^{15}N-labelled protein and $[L]_0$ is the total ligand concentration. The K_D and $\Delta\nu_{max}$ were used as fitting parameters. [^{1}H, ^{15}N]-HSQC spectra were processed and analysed using NMRView.[254] For chemical shift mapping, PDB

structures (PDB ID: 3D3B) are analysed in PyMol, the backbone amide residues have previously been assigned by the Rösch group in the Department of Biopolymers, University of Bayreuth. A threshold value of $\Delta\delta(^{1}H/^{15}N)$ = 0.04 ppm is considered to indicate that the corresponding amino acid is directly involved in the binding interaction.[178] The residues in the crystal structure (PDB ID: 3D3B) are then colour coded according to the $\Delta\delta_{norm}$: yellow (0.04 – 0.10 ppm), orange (0.10 – 0.20 ppm) and red (> 0.20 ppm. Structures are visualised using PyMol.

Chapter 8

Appendix

8.1. Appendix for chapter 2

8.1.1. ATAD2 Protein Expression and Purification

Parameter	Result
Sequence	MHHHHHHSSGVDLGTENLYFQSMQEEDTFRELRIFLRN VTHRLAIDKRFRVFTKPVDPDEVPDYVTVIKQPMDLSSVI SKIDLHKYLTVKDYLRDIDLICSNALEYNPDRDPGDRLIR HRACALRDTAYAIIKEELDEDFEQLCEEIQESR
No. of amino acids	151
Molecular Weight (Da)	18114
Theoretical pI	5.6
Extinction coefficient (M^{-1}. cm^{-1})	8940

Table 8.1: Calculated parameters of ATAD2 using ExPASy ProtParam online calculator. Extinction coefficient and theoretical pI used for concentration determination and buffer description, respectively.

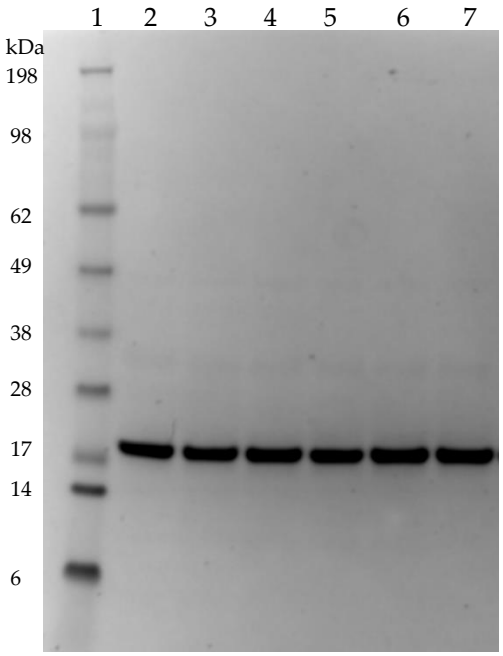


Figure 8.1: SDS-PAGE gel image of ATAD2 fractions 1 to 6 (Lanes 2-7). Fractions were pooled and aliquoted for storage at -80 °C. Lane 1: molecular weight marker (SeeBlue™ Plus2 Pre-stained Protein Standard, Invitrogen).

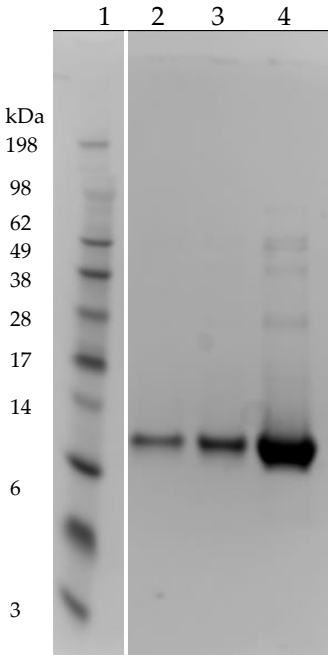


Figure 8.2: ^{15}N-labelled ATAD2 SDS-PAGE gel. Fractions after pooling at three different loading volumes. Lane 1: molecular weight ladder. Lane 2: 1 μg ^{15}N-ATAD2. Lane 3: 5 μg ^{15}N-ATAD2. Lane 4: 10 μg ^{15}N-ATAD2.

8.1.2. Fluorescent Labelling of ATAD2

Parameter	Value
Average absorbance 280 nm	0.047
Average absorbance 330 nm,	0.036
Average absorbance 650 nm	0.407
$A_{280\ corrected}$	0.031
$A_{330\ corrected}$	0.015
$A_{280\ stray\ light\ corrected}$	0.001
Total Protein Concentration	1.61 μM
Total Dye Concentration	16.2 μM
Ratio	1:10 (10 %)

Table 8.2: Absolute values of calculated absorbance measurements for Cy5 labelling of ATAD2.

8.1.3. UbE2L3 Protein Expression and Purification

Parameter	Result
Sequence	MASSIYINAASRRLMKELEEIRKCGMKNFRNIQVDEANL LTWQGLIVPDNPPYDKGAFRIEINFPAEYPFKPPKITFKTK IYHPNIDEKGQVCLPVISAENWKPATKTDQVIQSLIALVN DPQPEHPLRADLAEEYSKDRKKFCKNAEEFTKKYGEKRP VDHHHHHH
No. of amino acids	167
Molecular Weight (Da)	19433.2
Theoretical pI	8.67
Extinction coefficient (M^{-1}. cm^{-1})	20065

Table 8.3: Protein parameter results as calculated using the online computational tool ExPASy Protparam. The protein sequence is used as an input to calculate the molecular weight, theoretical pI and extinction coefficients.

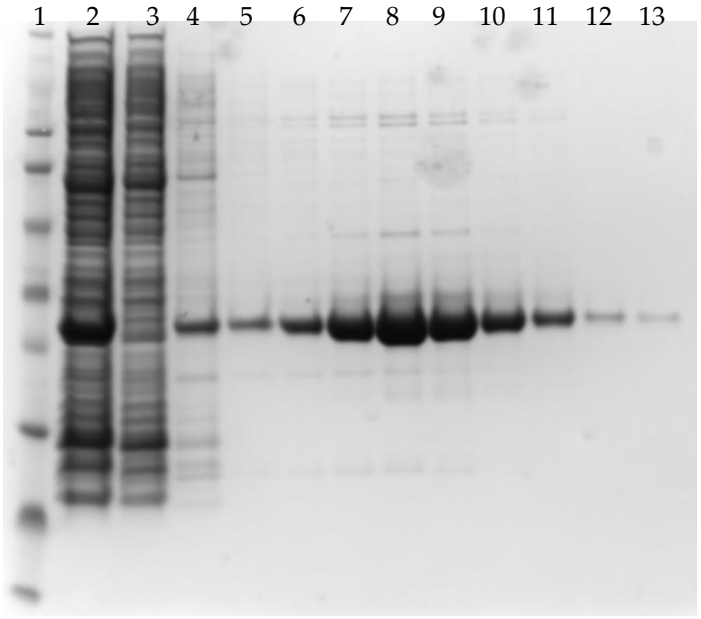


Figure 8.3: SDS-PAGE analysis of UbE2L3 purification. *Lane 1*: Molecular weight marker (SeeBlue Plus2 Pre-stained, ThermoFisher); *Lane 2*: Lysate; *Lane 3*: Column flow through; *Lane 4*: Column wash 1; *Lane 5:* Column wash 2; *Lane 6*: Fraction 1; *Lane 7*: Fraction 2; *Lane 8*: Fraction 3; *Lane 9*: Fraction 4; *Lane 10*: Fraction 5; *Lane 11*: Fraction 6; *Lane 12*: Fraction 7; *Lane 13*: Fraction 8.

8.1.4. Synthesis of Fmoc-ε-acetyl-*L*-lysine

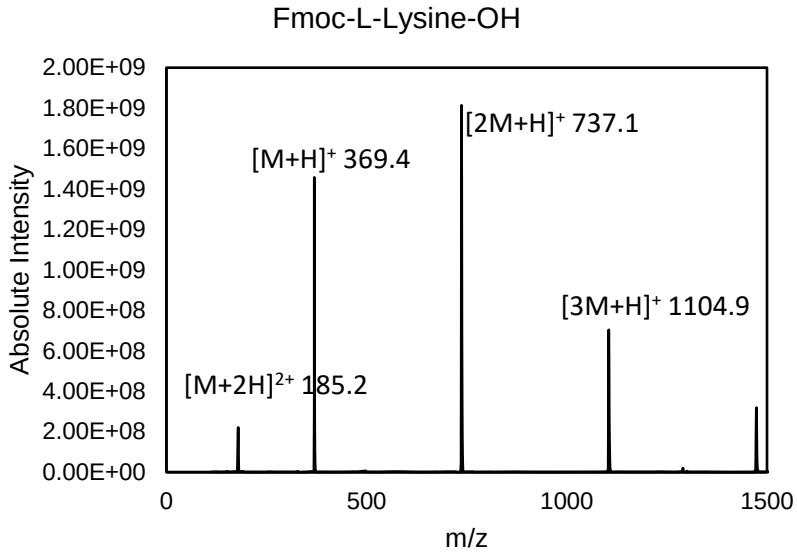


Figure 8.4: LCMS mass spectrum of N_{α}-Fmoc-*L*-lysine-OH. $[M+H]^{+}_{calc}$ 369.2; $[M+H]^{+}_{obs}$ 369.4.

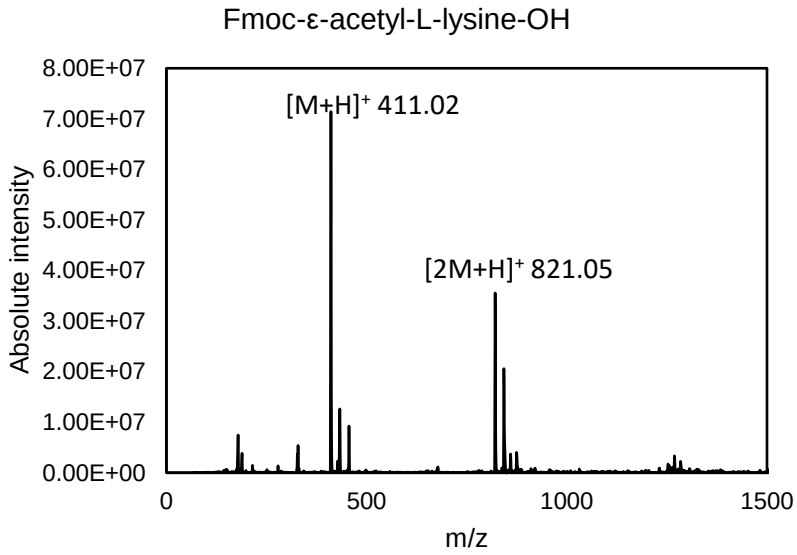


Figure 8.5: LCMS mass spectrum of Fmoc-ε-acetyl-*L*-lysine-OH. $[M+H]^{+}_{calc}$ 411.2; $[M+H]^{+}_{obs}$ 411.02.

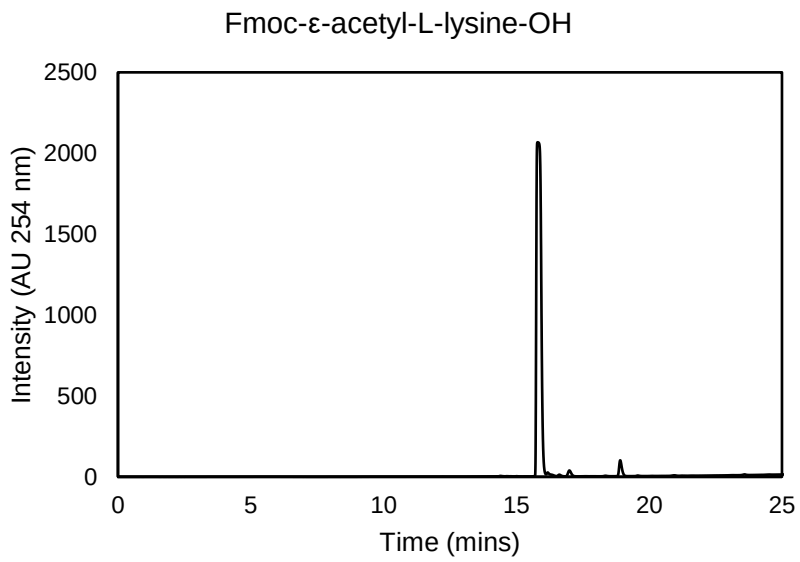


Figure 8.6: HPLC chromatogram Fmoc-N_{ε}-acetyl-*L*-lysine-OH. t_R 15.9 mins; > 95 % purity as measured by peak area.

8.1.5. Synthesis of Peptide H4K5

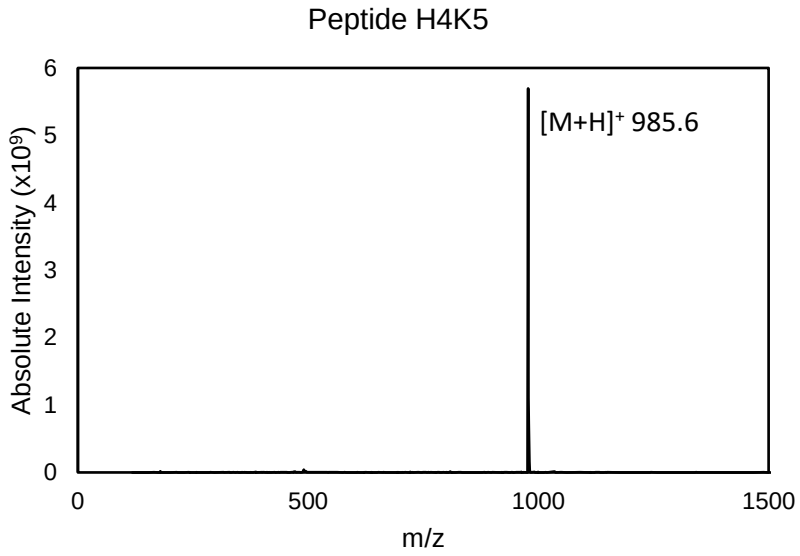


Figure 8.7: LCMS mass spectrum of peptide H4K5 (H_2N-RGK_{ac}GG-DOA-β(*m*CF_3)F-Pra). $[M+H]^+_{calc}$ 985.4; $[M+H]^+_{obs}$ 985.6.

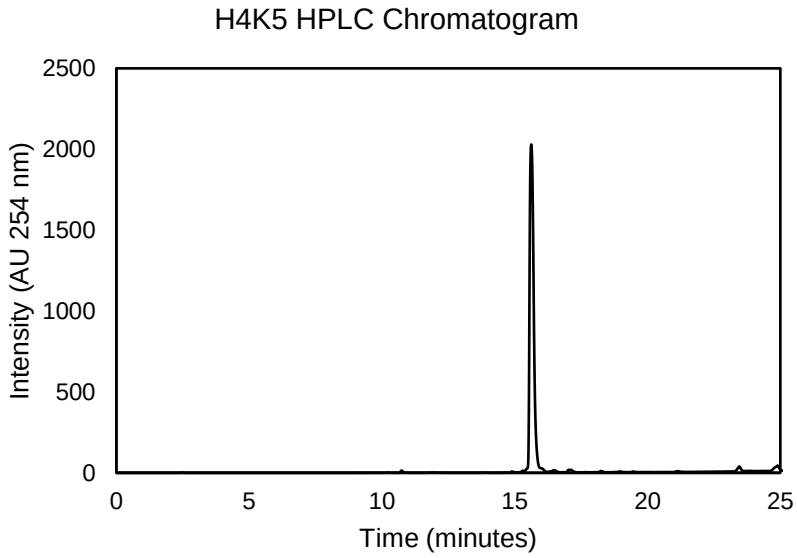


Figure 8.8: HPLC chromatogram of peptide H4K5. t_R = 15.6 mins. Absorbance is measured at 254 nm, and purity is > 95% (as measured by peak area).

8.1.6. Synthesis of Peptide H4

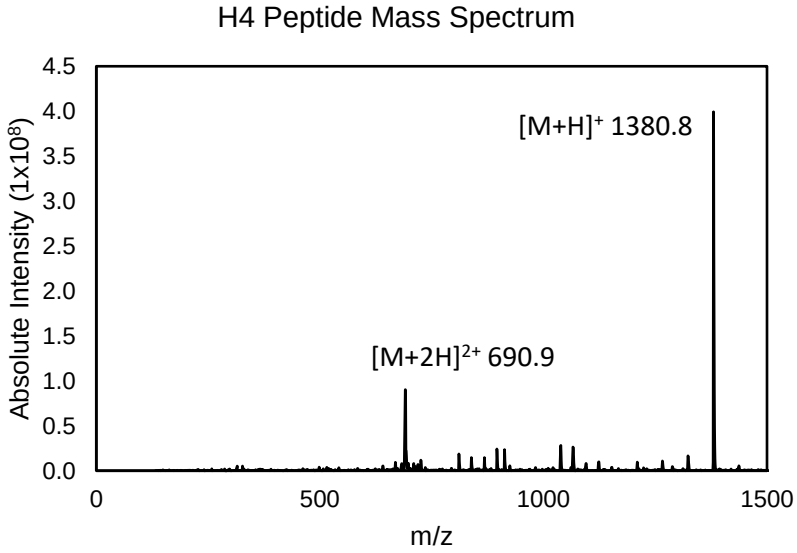


Figure 8.9: LCMS mass spectrum of peptide H4: $C_{59}H_{102}N_{20}O_{18}$ (H_2N-RGK$_{ac}$GGK$_{ac}$GLGK$_{ac}$G-DOA-PRA-OH). $[M+H]^+_{calc}$ 1378.8; $[M+H]^+_{obs}$ 1380.8.

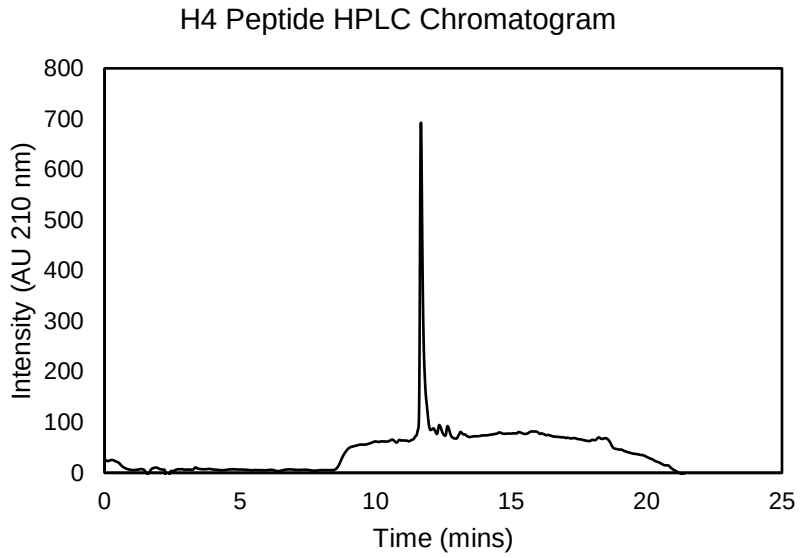


Figure 8.10: HPLC chromatogram of peptide H4. t_R = 11.7 mins. Absorbance is measured at 210 nm, and purity is > 90% (as measured by peak area).

8.1.7. Synthesis of Peptide hoA-(*m*CF_3)F-E-R and di/tripeptide truncates

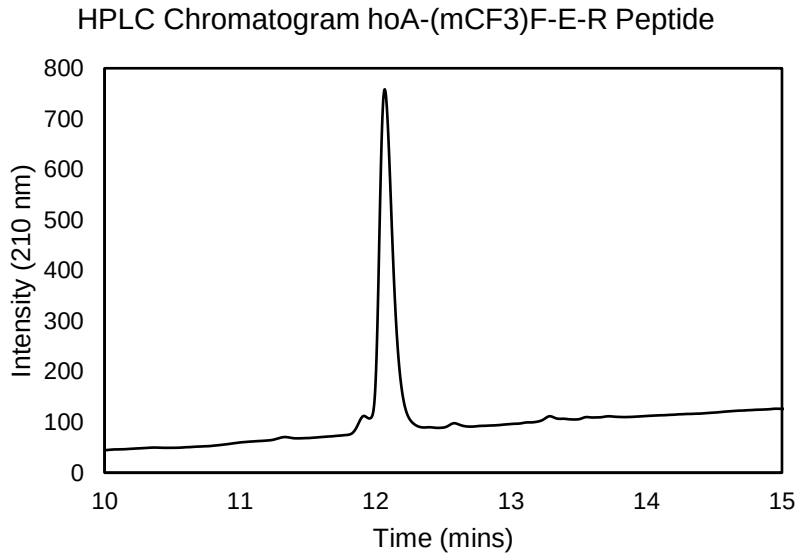


Figure 8.11: HPLC chromatogram of peptide H_2N-hoA-(mCF3)F-E-R-$CONH_2$. t_R = 12.1 mins, > 95 % purity as measured by peak area.

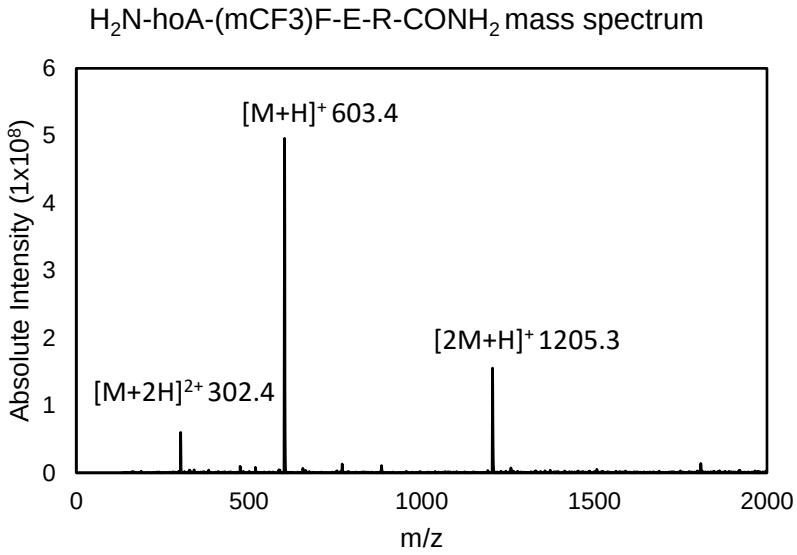


Figure 8.12: LCMS mass spectrum of peptide H_2N-hoA-(mCF3)F-E-R-$CONH_2$. $C_{25}H_{37}F_3N_8O_6$: $[M+H]^+_{calc}$ 603.3; $[M+H]^+_{obs}$ 603.4.

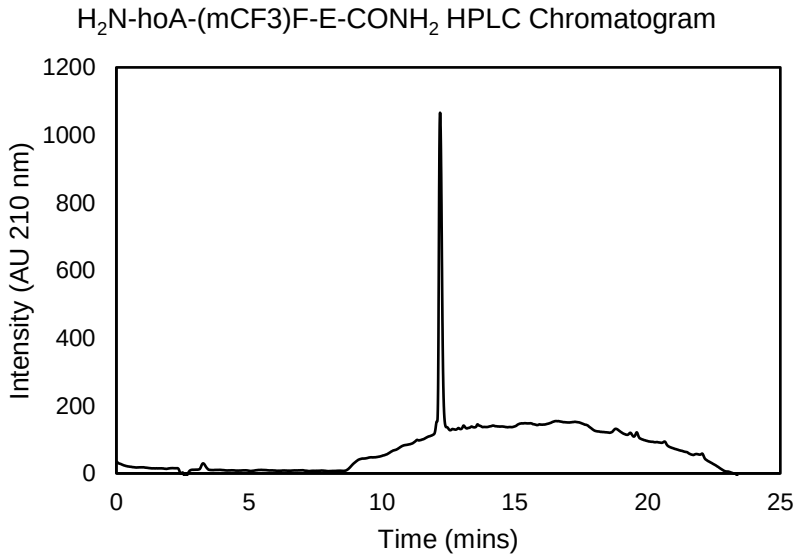


Figure 8.13: HPLC chromatogram of peptide H_2N-hoA-(mCF3)F-E-$CONH_2$. t_R = 12.2 mins, > 95 % purity as measured by peak area.

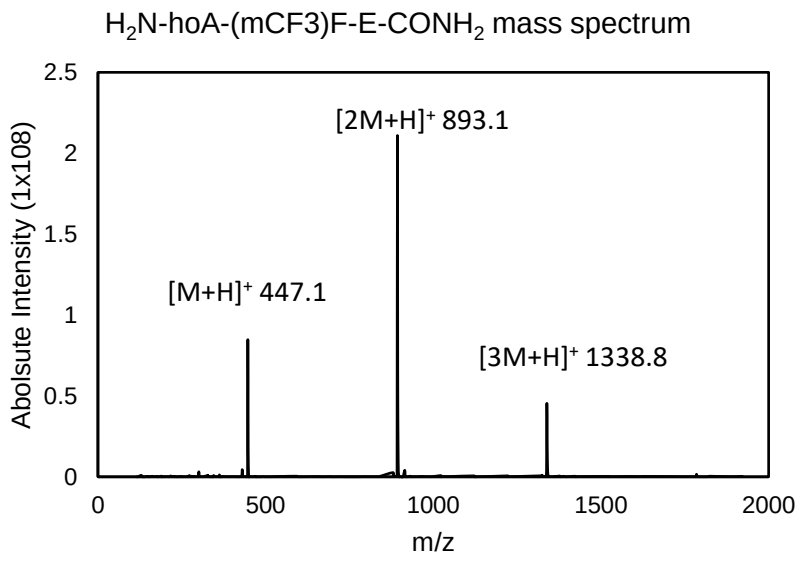


Figure 8.14: LCMS mass spectrum of peptide H_2N-hoA-(mCF3)F-E-$CONH_2$. $C_{19}H_{25}F_3N_4O_5$: $[M+H]^+_{calc}$ 447.2; $[M+H]^+_{obs}$ 447.1.

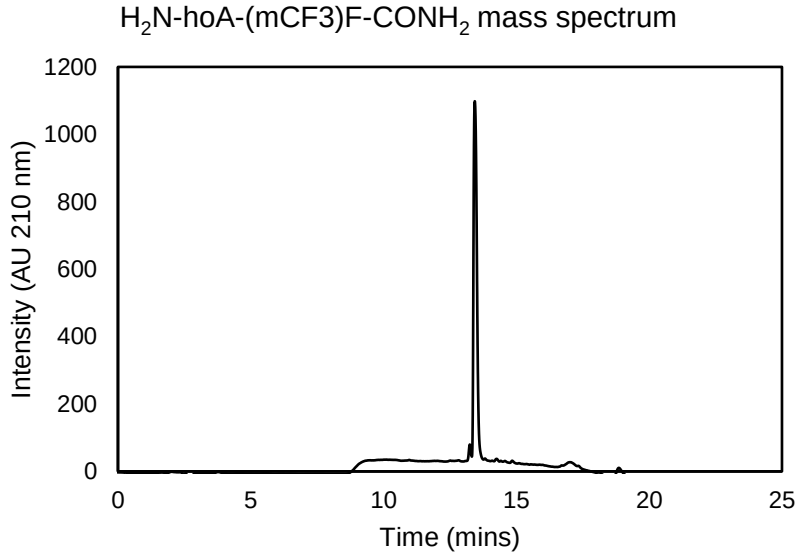


Figure 8.15: HPLC chromatogram of peptide H_2N-hoA-(mCF3)F-$CONH_2$. t_R = 13.5 mins, > 95 % purity as measured by peak area.

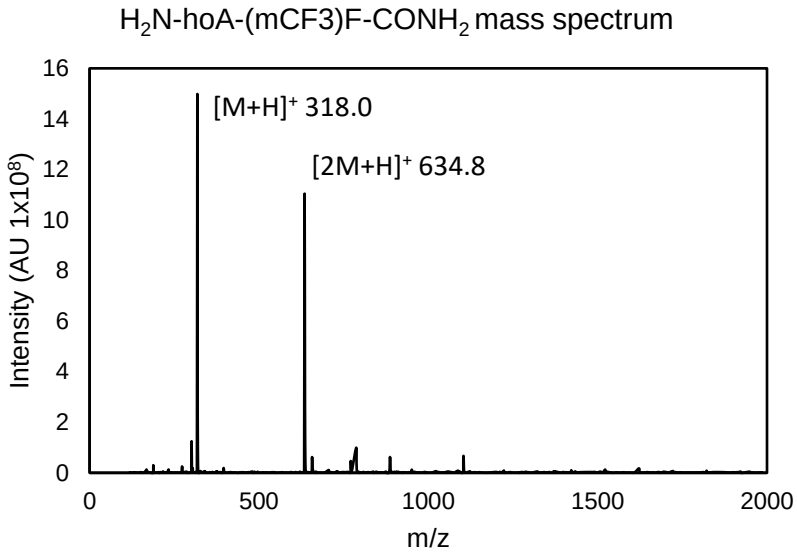


Figure 8.16: LCMS mass spectrum of peptide H_2N-hoA-(mCF3)F-$CONH_2$. $C_{14}H_{18}F_3N_3O_2$: $[M+H]^+_{calc}$ 318.1; $[M+H]^+_{obs}$ 318.0.

8.1.8. OBOC Purity analysis by TAMRA Concentration

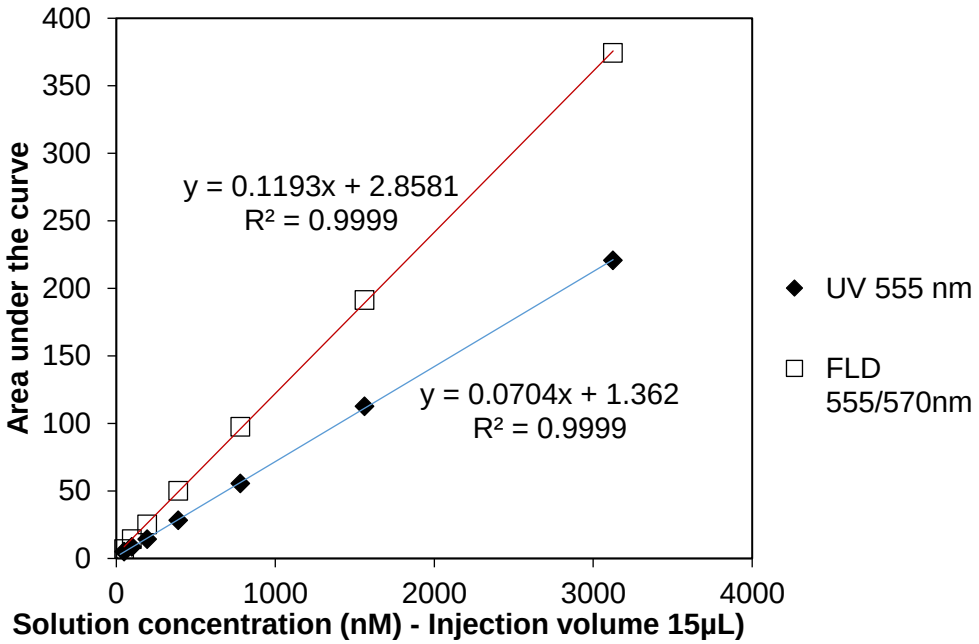
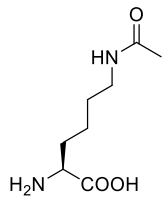
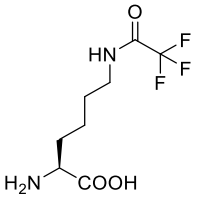


Figure 8.17: TMR-N_3 concentration curve. Achieved by Christophe Portal in the Auer group.

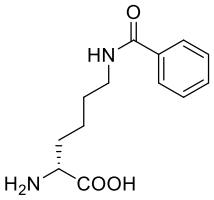
8.1.9. OBOC Lysine-derivative library compounds



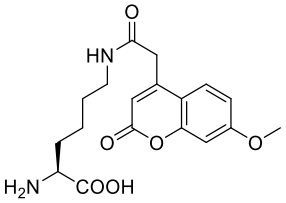
Lys-DEV-1



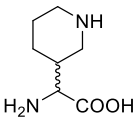
Lys-DEV-2



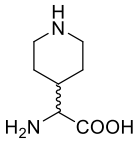
Lys-DEV-3



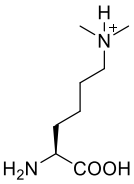
Lys-DEV-4



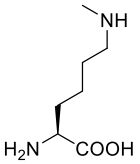
Lys-DEV-5



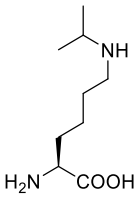
Lys-DEV-6



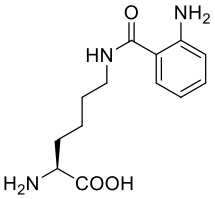
Lys-DEV-7



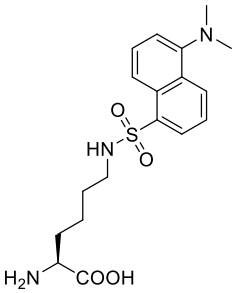
Lys-DEV-8



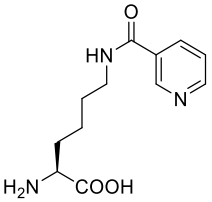
Lys-DEV-9



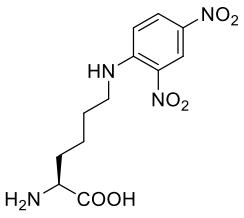
Lys-DEV-10



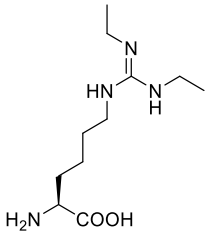
Lys-DEV-11



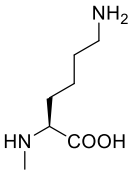
Lys-DEV-12



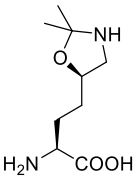
Lys-DEV-13



Lys-DEV-14



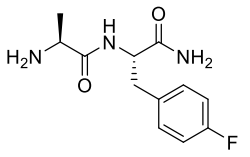
Lys-DEV-15



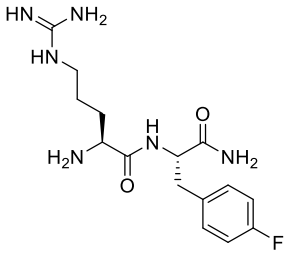
Lys-DEV-16

Table 8.4: Chemical structures of lysine derivatives used in synthesis of OBOC library. Lys-DEVs 7, 9, 13, 14 failed QC checks, therefore were not included in the final library.

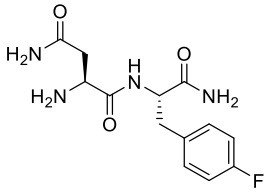
8.1.10. Fluorinated Dipeptides – Structures and QC data



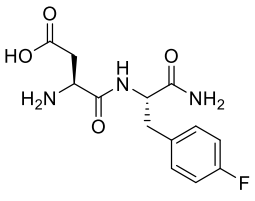
FDL01



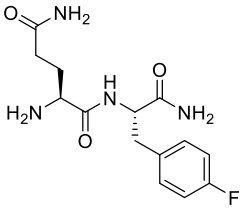
FDL02



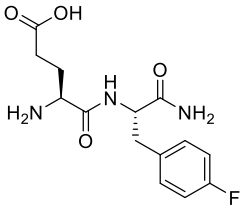
FDL03



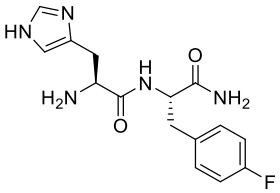
FDL04



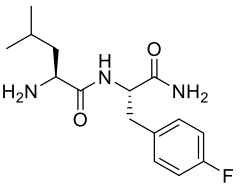
FDL05



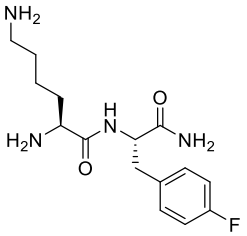
FDL06



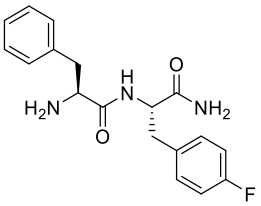
FDL07



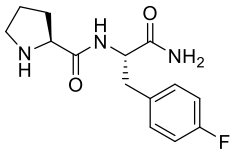
FDL08



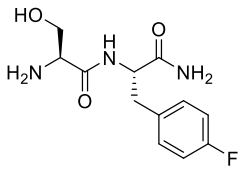
FDL09



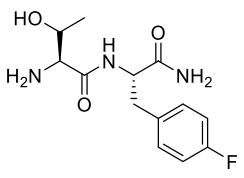
FDL10



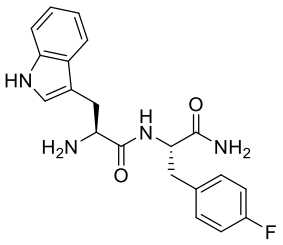
FDL11



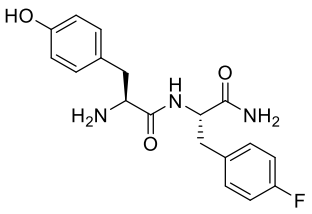
FDL12



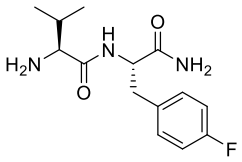
FDL13



FDL14



FDL15

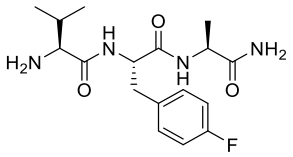


FDL16

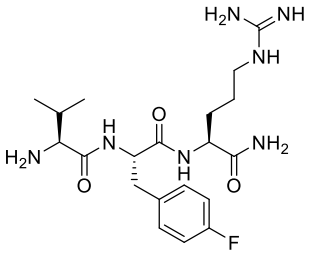
FDL01. A*p*F. MS (ESI): $[M+H]^{+}_{calc}$ 254.3; $[M+H]^{+}_{obs}$ 254.0. HPLC: t_R 10.3 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43768.4. **FDL02.** R*p*F. MS (ESI): $[M+H]^{+}_{calc}$ 339.4; $[M+H]^{+}_{obs}$ 339.2. HPLC: t_R 10.9 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43769.55. **FDL03.** N*p*F. MS (ESI): $[M+H]^{+}_{calc}$ 297.3; $[M+H]^{+}_{obs}$ 297.0. HPLC: t_R 10.7 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43787.93. **FDL04.** D*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 298.3; $[M+H]^{+}_{obs}$ 298.1. HPLC: t_R 10.8 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43784.05. **FDL05.** Q*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 311.3; $[M+H]^{+}_{obs}$ 311.0. HPLC: t_R 10.7 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43757.28. **FDL06.** E*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 312.2; $[M+H]^{+}_{obs}$ 312.0. HPLC: t_R 10.9 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43750.73. **FDL07.** H*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 320.3; $[M+H]^{+}_{obs}$ 320.2. HPLC: t_R 10.8 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43780.85. **FDL08.** L*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 296.4; $[M+H]^{+}_{obs}$ 295.99. HPLC: t_R 12.9 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43792.75. **FDL09.** K*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 311.4; $[M+H]^{+}_{obs}$ 311.2. HPLC: t_R 10.6 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43786.08. **FDL10.** F*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 330.4; $[M+H]^{+}_{obs}$ 330.0. HPLC: t_R 13.3 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43795.27. **FDL11.** P*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 280.3; $[M+H]^{+}_{obs}$ 280.1. HPLC: t_R 11.3 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43743.99. **FDL12.** S*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 270.3; $[M+H]^{+}_{obs}$ 270.0. HPLC: t_R 10.6 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43788.53. **FDL13.** T*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 284.3; $[M+H]^{+}_{obs}$ 283.9. HPLC: t_R 10.9 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43785.86. **FDL14.** W*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 369.4; $[M+H]^{+}_{obs}$ 369.0. HPLC: t_R 14.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43779.53. **FDL15.** Y*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 346.4; $[M+H]^{+}_{obs}$ 346.0. HPLC: t_R 12.2 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43779.61. **FDL16.** V*p*F. MS (ESI) $[M+H]^{+}_{calc}$ 282.3; $[M+H]^{+}_{obs}$ 282.0. HPLC: t_R 11.7 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43775.20.

FDL Pools. *FDL-Pool-1*: 10, 11, 13, 16. *FDL-Pool-2*: 02, 04, 08. *FDL-Pool-3*: 01, 07, 12. *FDL-Pool-4:* 03, 05, 14. *FDL-Pool-5*: 06, 09, 15.

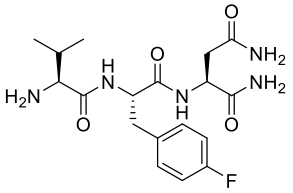
8.1.11. C-Terminal Extended Fluorine Tripeptides - Structures and QC data



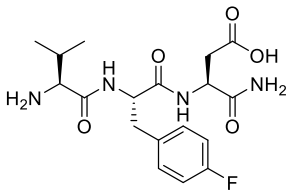
C-FTL01



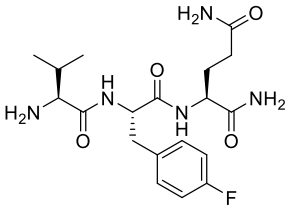
C-FTL02



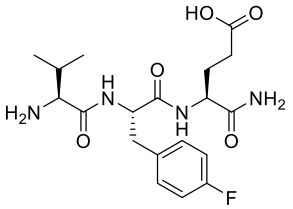
C-FTL03



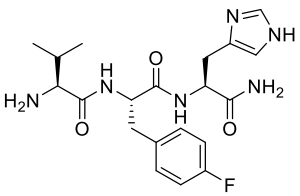
C-FTL04



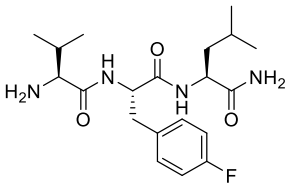
C-FTL05



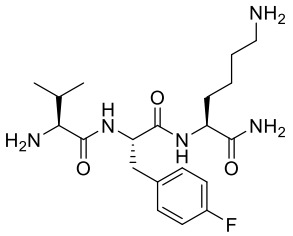
C-FTL06



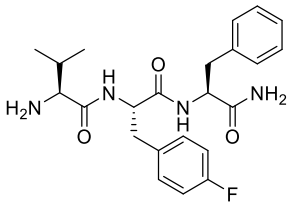
C-FTL07



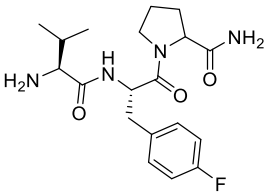
C-FTL08



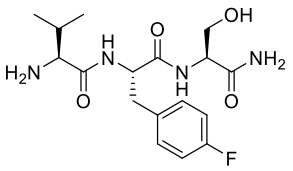
C-FTL09



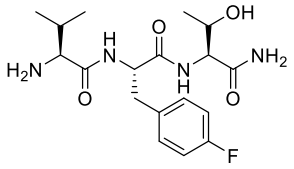
C-FTL10



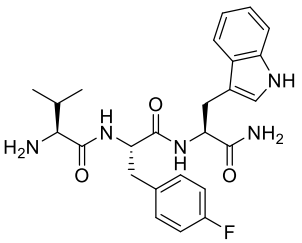
C-FTL11



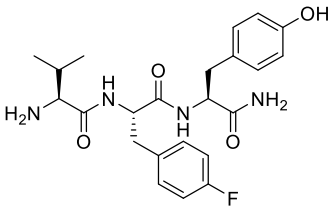
C-FTL12



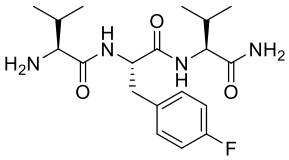
C-FTL13



C-FTL14



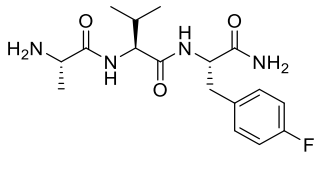
C-FTL15



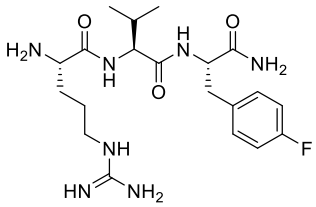
C-FTL16

C-FTL01. V*p*FA. MS (ESI): $[M+H]^{+}_{calc}$ 353.2; $[M+H]^{+}_{obs}$ 353.2. HPLC: t_R 11.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43688.12. **C-FTL02.** V*p*FR. MS (ESI): $[M+H]^{+}_{calc}$ 438.3; $[M+H]^{+}_{obs}$ 438.3. HPLC: t_R 11.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43657.09. **C-FTL03.** V*p*FN. MS (ESI): $[M+H]^{+}_{calc}$ 396.2; $[M+H]^{+}_{obs}$ 396.2. HPLC: t_R 11.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43668.39. **C-FTL04.** V*p*FD. MS (ESI): $[M+H]^{+}_{calc}$ 397.2; $[M+H]^{+}_{obs}$ 397.2. HPLC: t_R 11.3 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43718.09. **C-FTL05.** V*p*FQ. MS (ESI): $[M+H]^{+}_{calc}$ 410.2; $[M+H]^{+}_{obs}$ 410.6. HPLC: t_R 11.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43656.87. **C-FTL06.** V*p*FE. MS (ESI): $[M+H]^{+}_{calc}$ 411.2; $[M+H]^{+}_{obs}$ 411.2. HPLC: t_R 11.3 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43708.41. **C-FTL07.** V*p*FH. MS (ESI): $[M+H]^{+}_{calc}$ 419.2; $[M+H]^{+}_{obs}$ 419.3. HPLC: t_R 10.8 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43681.60. **C-FTL08.** V*p*FL. MS (ESI): $[M+H]^{+}_{calc}$ 395.2; $[M+H]^{+}_{obs}$ 395.2. HPLC: t_R 13.6 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43676.48. **C-FTL09.** V*p*FK. MS (ESI): $[M+H]^{+}_{calc}$ 410.2; $[M+H]^{+}_{obs}$ 410.3. HPLC: t_R 10.7 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43668.50. **C-FTL10.** V*p*FF. MS (ESI): $[M+H]^{+}_{calc}$ 429.2; $[M+H]^{+}_{obs}$ 429.2. HPLC: t_R 14.2 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43704.91. **C-FTL11.** V*p*FP. MS (ESI): $[M+H]^{+}_{calc}$ 379.2; $[M+H]^{+}_{obs}$ 379.2. HPLC: t_R 12.4 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43676.30. **C-FTL12.** V*p*FS. MS (ESI): $[M+H]^{+}_{calc}$ 369.2; $[M+H]^{+}_{obs}$ 369.2. HPLC: t_R 11.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43672.87. **C-FTL13.** V*p*FT. MS (ESI): $[M+H]^{+}_{calc}$ 383.2; $[M+H]^{+}_{obs}$ 383.2. HPLC: t_R 11.4 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43660.71. **C-FTL14.** V*p*FW. MS (ESI): $[M+H]^{+}_{calc}$ 468.2; $[M+H]^{+}_{obs}$ 468.3. HPLC: t_R 14.5 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43691.05. **C-FTL15.** V*p*FY. MS (ESI): $[M+H]^{+}_{calc}$ 445.2; $[M+H]^{+}_{obs}$ 445.2. HPLC: t_R 12.7 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43691.66. **C-FTL16.** V*p*FV. MS (ESI): $[M+H]^{+}_{calc}$ 381.2; $[M+H]^{+}_{obs}$ 381.2. HPLC: t_R 12.5 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43715.08.

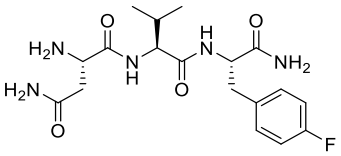
8.1.12. N-Terminal Extended Fluorine Tripeptides – Structures and QC data



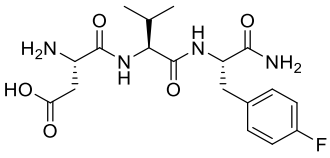
N-FTL01



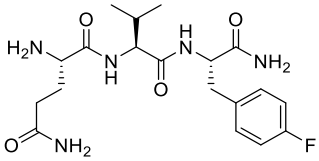
N-FTL02



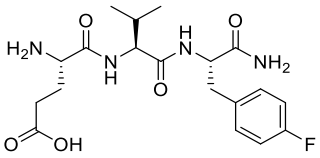
N-FTL03



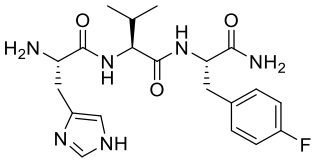
N-FTL04



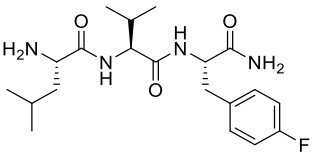
N-FTL05



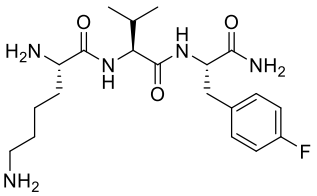
N-FTL06



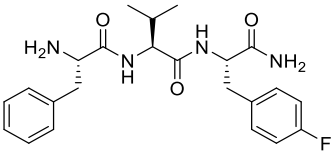
N-FTL07



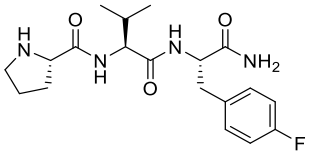
N-FTL08



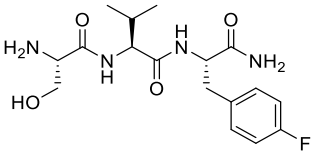
N-FTL09



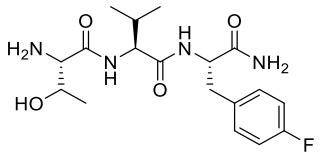
N-FTL10



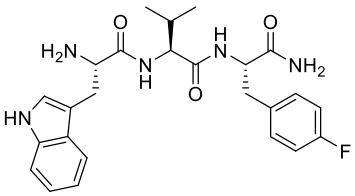
N-FTL11



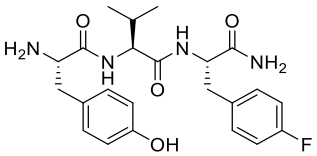
N-FTL12



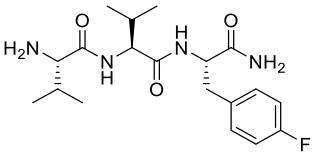
N-FTL13



N-FTL14



N-FTL15



N-FTL16

QC Data for all N-FTL library compounds

N-FTL01. AV*p*F. MS (ESI): $[M+H]^+_{calc}$ 353.2; $[M+H]^+_{obs}$ 353.3. HPLC: t_R 12.4 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43857.32. **N-FTL02.** RV*p*F. MS (ESI): $[M+H]^+_{calc}$ 438.3; $[M+H]^+_{obs}$ 438.3. HPLC: t_R 12.3 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43846.06. **N-FTL03.** NV*p*F. MS (ESI): $[M+H]^+_{calc}$ 396.2; $[M+H]^+_{obs}$ 396.2. HPLC: t_R 12.9 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43869.67. **N-FTL04.** DV*p*F. MS (ESI): $[M+H]^+_{calc}$ 397.2; $[M+H]^+_{obs}$ 397.2. HPLC: t_R 12.9 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43850.50. **N-FTL05.** QV*p*F. MS (ESI): $[M+H]^+_{calc}$ 410.2; $[M+H]^+_{obs}$ 410.2. HPLC: t_R 12.4 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43801.93. **N-FTL06.** EV*p*F. MS (ESI): $[M+H]^+_{calc}$ 411.2; $[M+H]^+_{obs}$ 411.2. HPLC: t_R 12.5 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43785.33. **N-FTL07.** HV*p*F. MS (ESI): $[M+H]^+_{calc}$ 419.2; $[M+H]^+_{obs}$ 419.3. HPLC: t_R 12.2 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43841.73. **N-FTL08.** LV*p*F. MS (ESI): $[M+H]^+_{calc}$ 395.2; $[M+H]^+_{obs}$ 395.2. HPLC: t_R 14.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43761.46. **N-FTL09.** KV*p*F. MS (ESI): $[M+H]^+_{calc}$ 410.2; $[M+H]^+_{obs}$ 410.3. HPLC: t_R 12.1 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43840.07. **N-FTL10.** FV*p*F. MS (ESI): $[M+H]^+_{calc}$ 429.2; $[M+H]^+_{obs}$ 429.3. HPLC: t_R 14.5 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43814.32. **N-FTL11.** PV*p*F. MS (ESI): $[M+H]^+_{calc}$ 379.2; $[M+H]^+_{obs}$ 379.3. HPLC: t_R 13.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43888.68. **N-FTL12.** SV*p*F. MS (ESI): $[M+H]^+_{calc}$ 369.2; $[M+H]^+_{obs}$ 369.2. HPLC: t_R 12.8 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43863.87. **N-FTL13.** TV*p*F. MS (ESI): $[M+H]^+_{calc}$ 383.2; $[M+H]^+_{obs}$ 383.2. HPLC: t_R 12.6 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43889.70. **N-FTL14.** WV*p*F. MS (ESI): $[M+H]^+_{calc}$ 468.2; $[M+H]^+_{obs}$ 468.2. HPLC: t_R 15.0 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43803.25. **N-FTL15.** YV*p*F. MS (ESI): $[M+H]^+_{calc}$ 445.2; $[M+H]^+_{obs}$ 445.2. HPLC: t_R 13.6 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43816.58. **N-FTL16.** VV*p*F. MS (ESI): $[M+H]^+_{calc}$ 381.2; $[M+H]^+_{obs}$ 381.2. HPLC: t_R 13.1 mins (> 95 %). NMR ^{19}F (376 MHz, 10 % D_2O) δ Hz -43873.69.

8.1.13. Fluorine Dipeptide Library Screening against UbE2L3

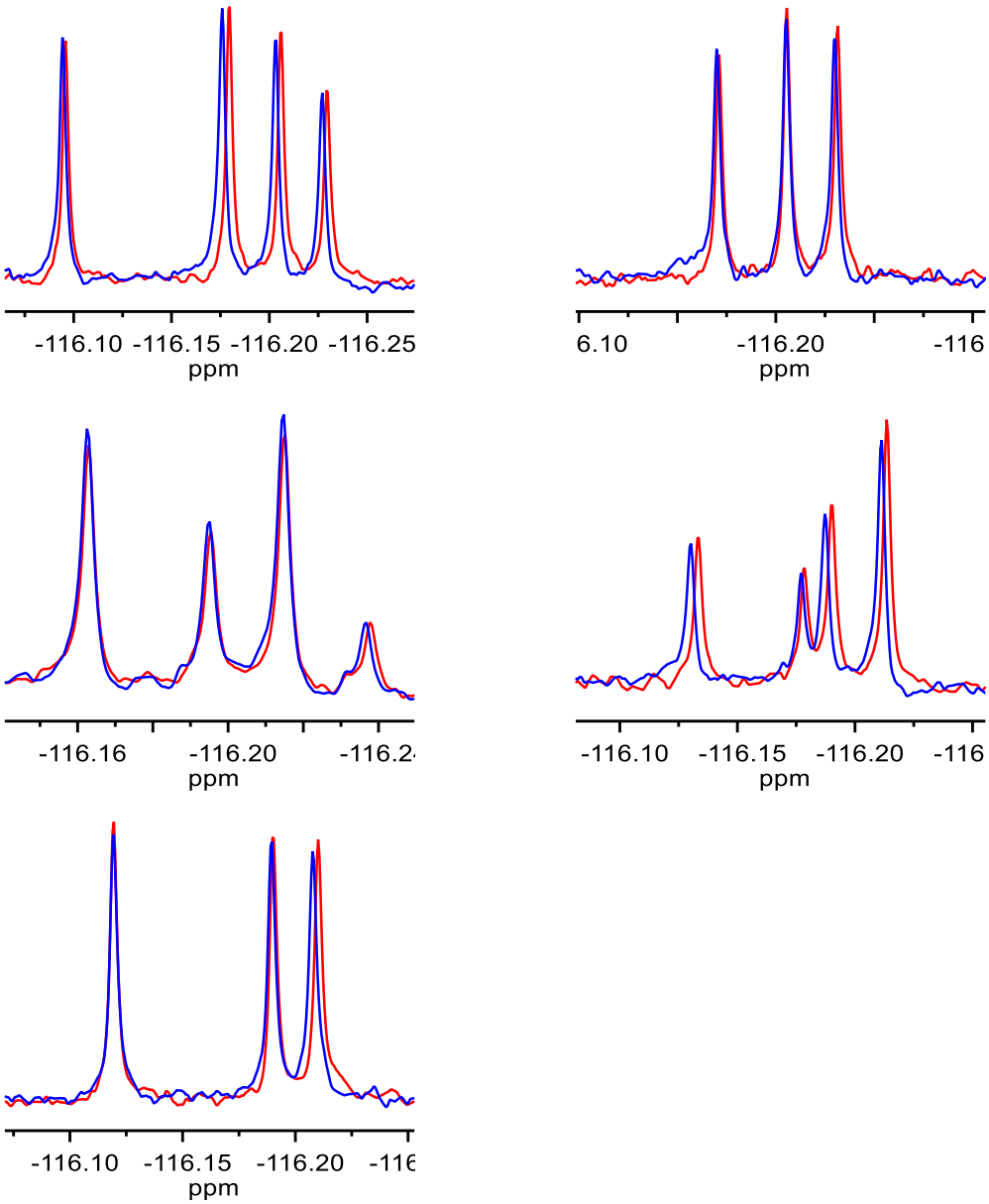


Figure 8.18: 1D ^{19}F NMR spectra of fluorinated dipeptides. Superposition of dipeptides (100 uM) in absence (red) and presence (blue) of 10 μM UbE2L3. Top left: FDL-Pool-1. Top right: FDL-Pool-2. Middle left: FDL-Pool-3. Middle right: FDL-Pool-4. Bottom left: FDL-Pool-5.

8.1.14. C-terminal Extended Fluorine Tripeptide Library (C-FTL) Screening

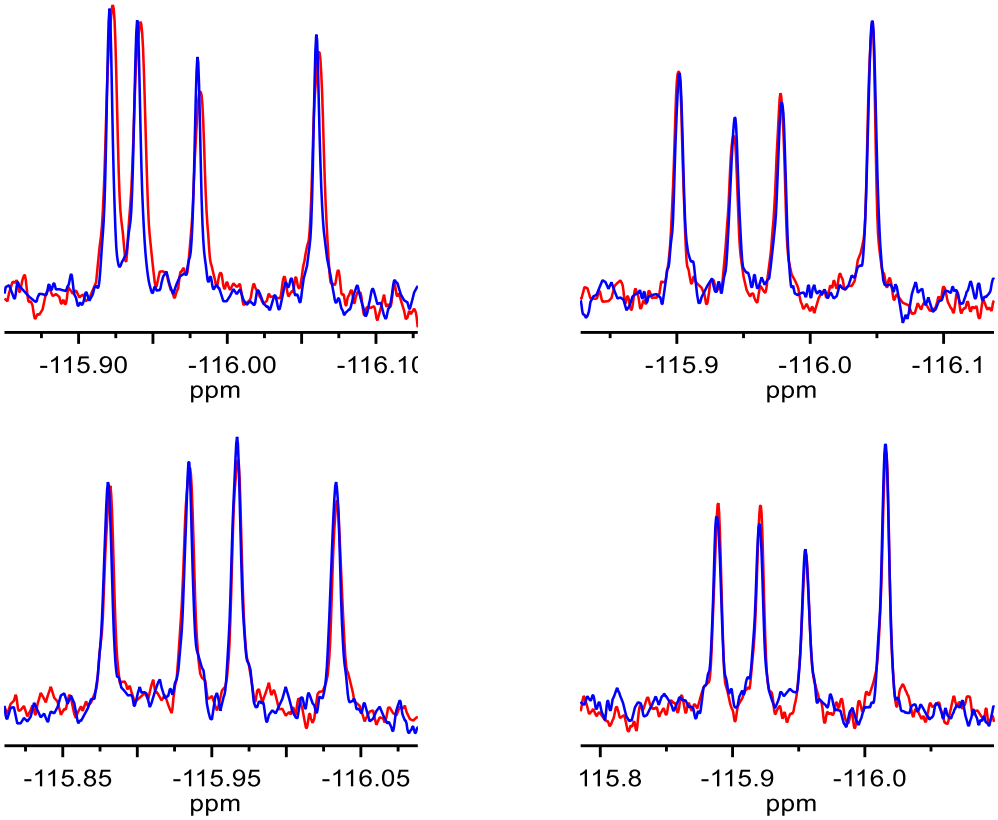


Figure 8.19: 1D ^{19}F NMR spectra of C-terminal extended fluorinated tripeptides (C-FTL). Superposition of tripeptides (100 μM) in absence (red) and presence (blue) of 10 μM UbE2L3. Top left: C-FTL-Pool-1. Top right: C-FTL-Pool-2. Bottom left: C-FTL-Pool-3. Bottom right: C-FTL-Pool-4.

8.1.15. N-terminal Extended Fluorine Tripeptide Library (N-FTL) Screening

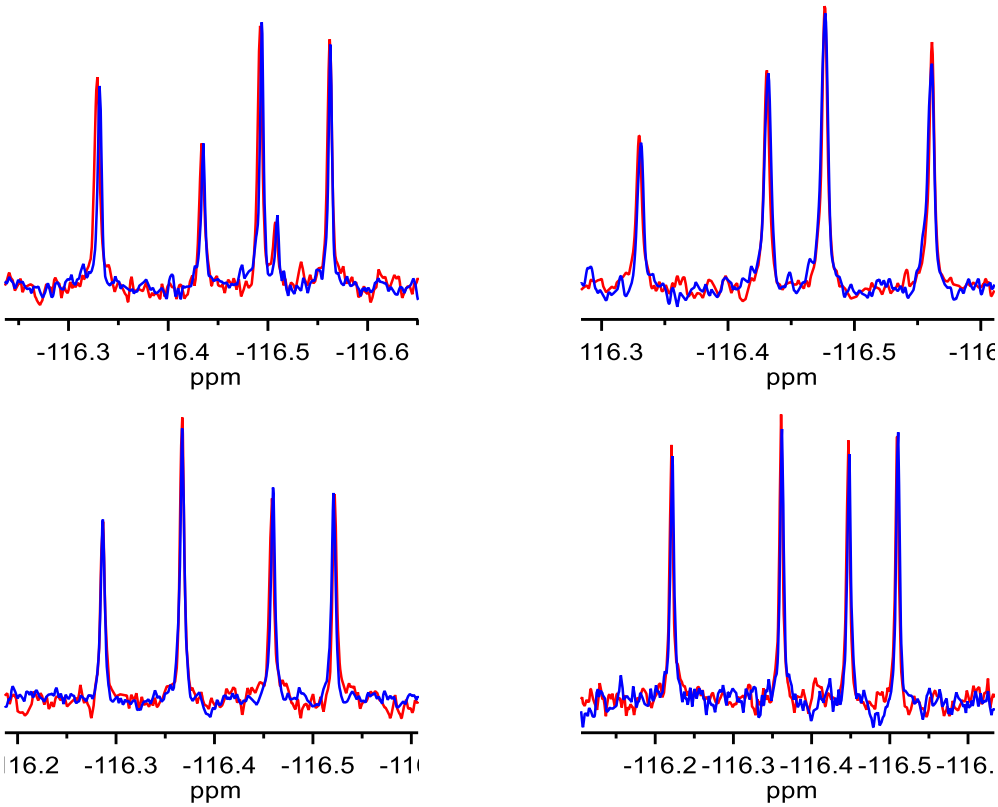
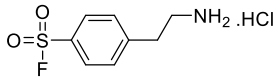


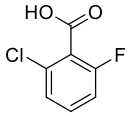
Figure 8.20: 1D ^{19}F NMR spectra of N-terminal extended fluorinated tripeptides (N-FTL). Superposition of tripeptides (100 μM) in absence (red) and presence (blue) of 10 μM UbE2L3. Top left: N-FTL-Pool-1. Top right: N-FTL-Pool-2. Bottom left: N-FTL-Pool-3. Bottom right: N-FTL-Pool-4.

8.2. Appendix for Chapter 3

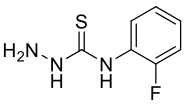
8.2.1. Recognised Active Fragment Library (RAFL) Compound Structures



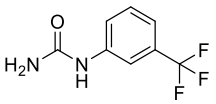
RAFL01



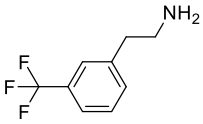
RAFL02



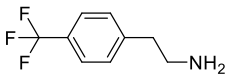
RAFL03



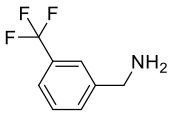
RAFL04



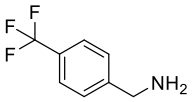
RAFL05



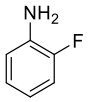
RAFL06



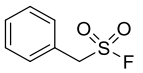
RAFL07



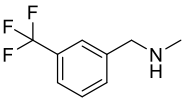
RAFL08



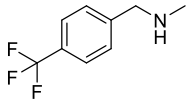
RAFL09



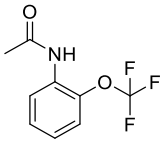
RAFL10



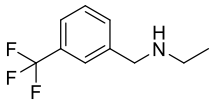
RAFL11



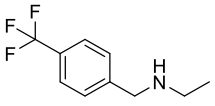
RAFL12



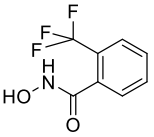
RAFL13



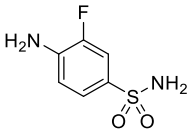
RAFL14



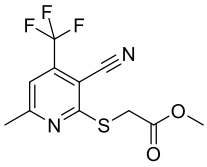
RAFL15



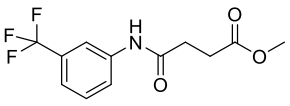
RAFL16



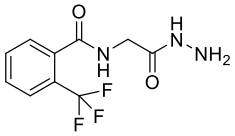
RAFL17



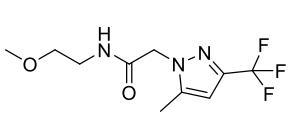
RAFL18



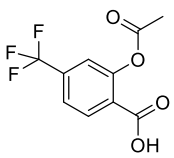
RAFL19



RAFL20

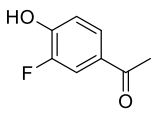


RAFL21

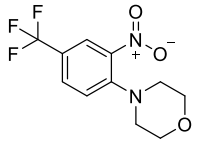


RAFL22

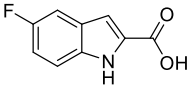
8.2.2. Clustered Fragment Library (CFL) Compound Structures



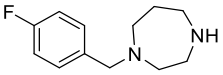
CFL01



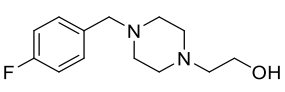
CFL02



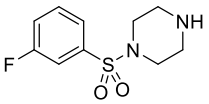
CFL03



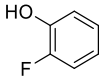
CFL04



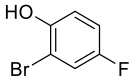
CFL05



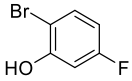
CFL06



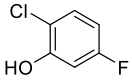
CFL07



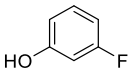
CFL08



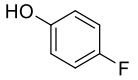
CFL09



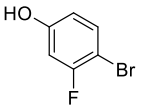
CFL10



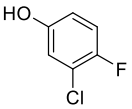
CFL11



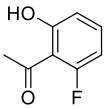
CFL12



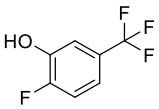
CFL13



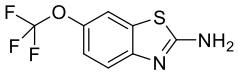
CFL14



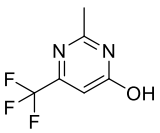
CFL15



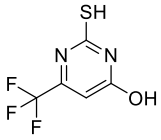
CFL16



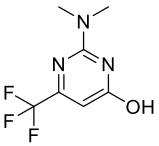
CFL17



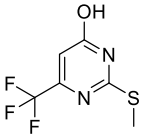
CFL18



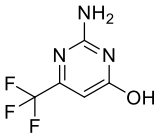
CFL19



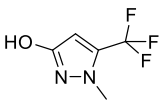
CFL20



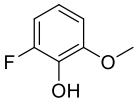
CFL21



CFL22

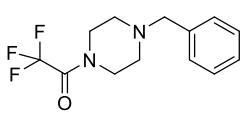


CFL23

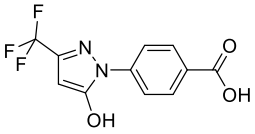


CFL24

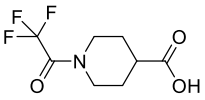
8.2.3. Directed Diversity Fragment Library (DDFL) Compound Structures



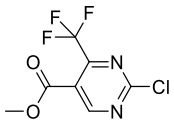
DDFL01



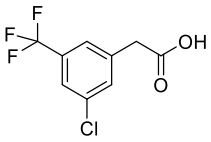
DDFL02



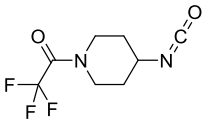
DDFL03



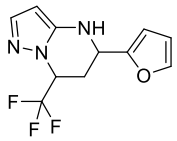
DDFL04



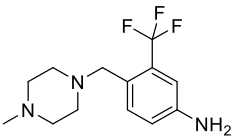
DDFL05



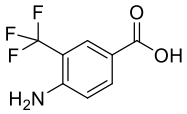
DDFL06



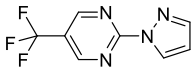
DDFL07



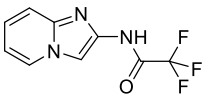
DDFL08



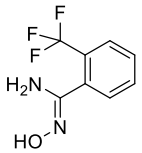
DDFL09



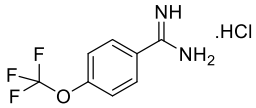
DDFL10



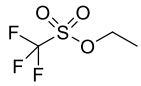
DDFL11



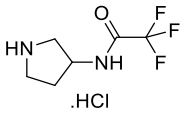
DDFL12



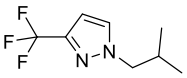
DDFL13



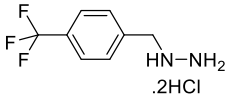
DDFL14



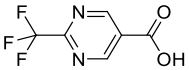
DDFL15



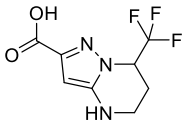
DDFL16



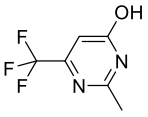
DDFL17



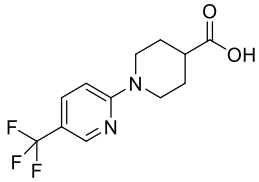
DDFL18



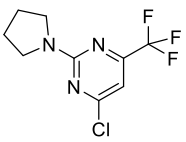
DDFL19



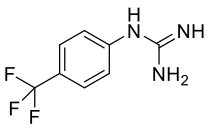
DDFL20



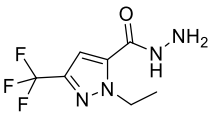
DDFL21



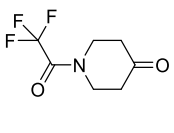
DDFL22



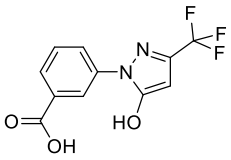
DDFL23



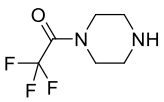
DDFL24



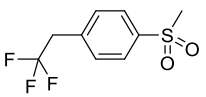
DDFL25



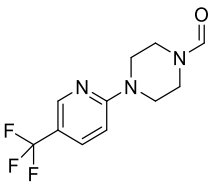
DDFL26



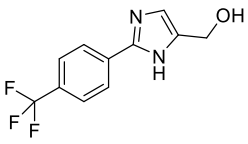
DDFL27



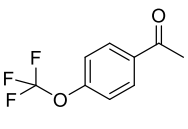
DDFL28



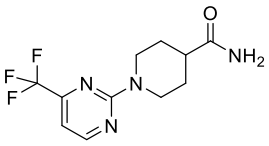
DDFL29



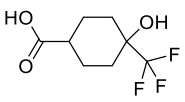
DDFL30



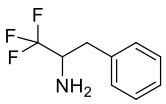
DDFL31



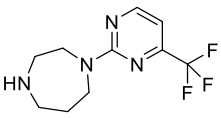
DDFL32



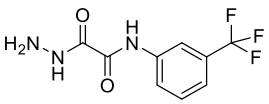
DDFL33



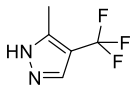
DDFL34



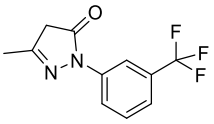
DDFL35



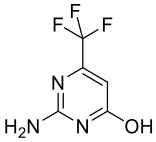
DDFL36



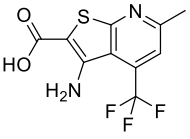
DDFL37



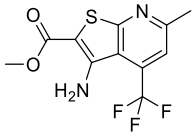
DDFL38



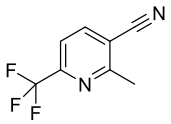
DDFL39



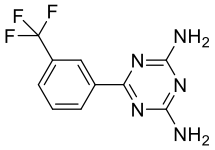
DDFL40



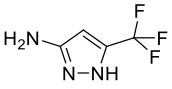
DDFL41



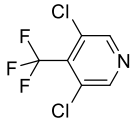
DDFL42

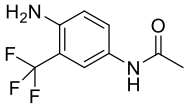


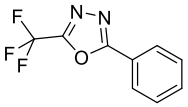
DDFL43

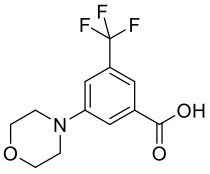


DDFL44

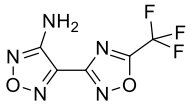
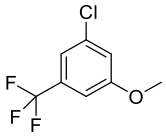
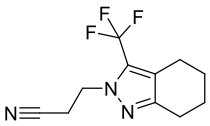
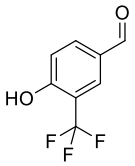




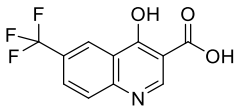
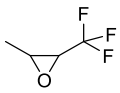
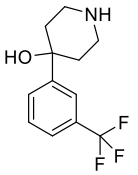
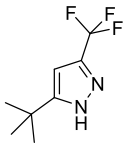




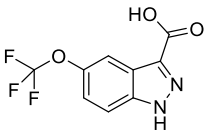
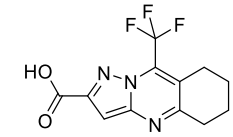
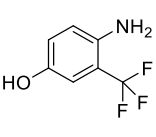
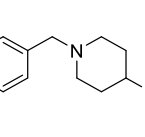
DDFL45 **DDFL46** **DDFL47** **DDFL48**

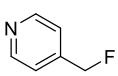
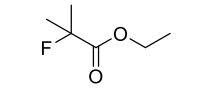
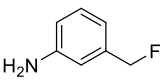
DDFL49 **DDFL50** **DDFL51** **DDFL52**

DDFL53 **DDFL54** **DDFL55** **DDFL56**

DDFL57 **DDFL58** **DDFL59** **DDFL60**

DDFL61 **DDFL62** **DDFL63**

8.3. Appendix for Chapter 4

8.3.1. Derivation of Equation 4.8

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$$

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[P]_f[L]_f}{[PL]}$$

$$[P]_f = [P]_0 - [PL]$$

$$[L]_f = [L]_0 - [PL]$$

$$K_D = \frac{([P]_0 - [PL])([L]_0 - [PL])}{[PL]}$$

$$K_D[PL] = [L]_0[P]_0 - [L]_0[PL] - [P]_0[PL] + [PL]^2$$

If $[P]_0 << [L]_0$, then $[PL][P]_0 + [PL]^2$ can be neglected:

$$K_D[PL] = [L]_0[P]_0 - [L]_0[PL]$$

$$[PL](K_D + [L]_0) = [L]_0[P]_0$$

$$\frac{[PL]}{[L]_0}(K_D + [L]_0) = [P]_0$$

The fraction of ligand bound, p_{Lb} is equal to:

$$\frac{[PL]}{[L]_0} = p_{Lb}$$

Therefore:

$$p_{Lb}(K_D + [L]_0) = [P]_0$$

And

$$p_{Lb} = \frac{[P]_0}{K_D + [L]_0}$$

Equations reproduced from Supplementary Data, Jordan *et al*.[153]

8.3.2. Derivation of Equation 4.10

$$p_{Lb,1}(K_D + [L]_1) = [P]_0$$

$$p_{Lb,2}(K_D + [L]_2) = [P]_0$$

$$\frac{p_{Lb,1}}{p_{Lb,2}} = \frac{(K_D + [L]_2)}{(K_D + [L]_1)}$$

Also, from equation 4.9:

$$p_{Lb,1} = \frac{(\nu_{1,obs} - \nu_{free})}{\Delta\nu^{app}} \text{ and } p_{Lb,2} = \frac{(\nu_{2,obs} - \nu_{free})}{\Delta\nu^{app}}$$

Therefore:

$$\frac{p_{Lb,1}}{p_{Lb,2}} = \frac{(\nu_{1,obs} - \nu_{free})}{(\nu_{2,obs} - \nu_{free})} = \gamma$$

And:

$$\gamma = \frac{(K_D + [L]_2)}{(K_D + [L]_1)}$$

Rearrange for K_D:

$$K_D = \frac{\gamma[L]_1 - [L]_2}{(1 - \gamma)}$$

Equations reproduced from Supplementary Data, Jordan *et al*.[45]

8.3.3. Recognised Active Fragment Library (RAFL) QC Data

~~**RAFL01.**~~ Formula: $C_8H_{10}FNO_2S$ (HCl). M_w 203.23 Da. MS (ESI): $[M+H]^+_{calc}$ 204.23; $[M+H]^+_{obs}$ N/A. HPLC: t_R 13.1 mins; > 95% purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **RAFL02.** Formula: $C_7H_4ClFO_2$. M_w 174.56 Da. MS (ESI): $[M+H]^+_{calc}$ 175.56; $[M+H]^+_{obs}$ N/A. HPLC: t_R 14.5 mins; > 95% purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -116.44. ~~**RAFL03.**~~ Formula: $C_7H_8FN_3S$. M_w 185.22 Da. MS (ESI): $[M+H]^+_{calc}$ 186.22 ; $[M+H]^+_{obs}$ N/A. HPLC: t_R N/A; purity N/A. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **RAFL04.** Formula: $C_8H_7F_3N_2O$. M_w 204.15 Da. MS (ESI): $[M+H]^+_{calc}$ 205.15; $[M+H]^+_{obs}$ 205.04. HPLC: t_R 15.6 mins; > 95% purity. NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.36. **RAFL05.** Formula: $C_9H_{10}F_3N$. M_w 189.18 Da. MS (ESI): $[M+H]^+_{calc}$ 190.18; $[M+H]^+_{obs}$ 189.97. HPLC: t_R 14.2 mins; > 95% purity. NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.23. **RAFL06.** Formula: $C_9H_{10}F_3N$. M_w 189.18 Da**.** MS (ESI): $[M+H]^+_{calc}$ 190.18; $[M+H]^+_{obs}$ 189.96**.** HPLC: t_R 14.3 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.06**. RAFL07.** Formula: $C_8H_8F_3N$. M_w 175.15 Da**.** MS (ESI): $[M+H]^+_{calc}$ 176.15; $[M+H]^+_{obs}$ 175.94**.** HPLC: t_R 13.0 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.38. **RAFL08.** Formula: $C_8H_8F_3N$. M_w 175.15 Da**.** MS (ESI): $[M+H]^+_{calc}$ 176.15; $[M+H]^+_{obs}$ 175.94**.** HPLC: t_R 13.2 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.38**.** ~~**RAFL09.**~~ Formula: C_6H_6FN. M_w 111.12 Da**.** MS (ESI): $[M+H]^+_{calc}$ 112.12; $[M+H]^+_{obs}$ N/A**.** HPLC: t_R 6.0 mins; > 95% purity**.**

NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm N/A. **~~RAFL10~~.** Formula: $C_7H_7FO_2S$. M_w 174.19 Da**.** MS (ESI): $[M+H]^+_{calc}$ 175.19; $[M+H]^+_{obs}$ N/A**.** HPLC: t_R 17.9 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm N/A. **RAFL11.** Formula: $C_9H_{10}F_3N$. M_w 189.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 190.1; $[M+H]^+_{obs}$ 190.0**.** HPLC: t_R 13.4 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.41. **RAFL12.** Formula: $C_9H_{10}F_3N$. M_w 189.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 190.1; $[M+H]^+_{obs}$ 190.0**.** HPLC: t_R 13.6 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.46. **RAFL13.** Formula: $C_9H_8F_3NO_2$. M_w 219.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 220.1; $[M+H]^+_{obs}$ 220.0**.** HPLC: t_R 16.5 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -57.94. **RAFL14.** Formula: $C_{10}H_{12}F_3N$. M_w 203.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 203.1; $[M+H]^+_{obs}$ N/A. HPLC: t_R 10.4 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.40. **~~RAFL15~~.** Formula: $C_{10}H_{12}F_3N$. M_w 203.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 203.1; $[M+H]^+_{obs}$ N/A. HPLC: t_R N/A**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm N/A. **RAFL16.** Formula: $C_8H_6F_3NO_2$. M_w 205.1 Da. MS (ESI):): $[M+H]^+_{calc}$ 206.0; $[M+H]^+_{obs}$ N/A. HPLC: t_R 11.2 mins; > 95 % purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -59.20. **~~RAFL17~~.** Formula: $C_6H_7FN_2O_2S$. M_w 190.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 191.0; $[M+H]^+_{obs}$ N/A. HPLC: t_R 6.2 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm N/A. **RAFL18.** Formula: $C_{11}H_9F_3N_2O_2S$. M_w 290.3 Da**.** MS (ESI): $[M+H]^+_{calc}$ 291.0; $[M+H]^+_{obs}$ 291.0. HPLC: t_R 20.0 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -63.97. **RAFL19.** Formula: $C_{12}H_{12}F_3NO_3$. M_w 275.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 276.1; $[M+H]^+_{obs}$ 276.1. HPLC: t_R 18.2 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.43. **RAFL20.** Formula: $C_{10}H_{10}F_3N_3O_2$. M_w 261.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 261.1; $[M+H]^+_{obs}$ N/A. HPLC: t_R 17.2 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -62.45. **RAFL21.** Formula: $C_{10}H_{14}F_3N_3O_2$. M_w 265.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 265.1; $[M+H]^+_{obs}$ N/A. HPLC: t_R 11.9 mins; > 95% purity**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm -58.95. **~~RAFL22~~.** Formula: $C_{10}H_7F_3O_4$. M_w 248.2 Da**.** MS (ESI): $[M+H]^+_{calc}$ 248.0; $[M+H]^+_{obs}$ N/A. HPLC: t_R N/A**.** NMR: ^{19}F (376 MHz, 10% D_2O) δ ppm N/A.

Scored out compounds did not pass QC checks, therefore were not included in the final screening pools. **RAFL Fragment Pools.** *RAFL-Pool-1*: 02, 07, 16, 19, 20. *RAFL-Pool-2*: 05, 08, 11, 18, 21. *RAFL-Pool-3*: 04, 06, 12, 13, 14.

8.3.4. Clustered Fragment Library (CFL) QC Data

CFL01. Formula: $C_8H_7FO_2$. M_w 154.14 Da. HPLC: t_R 13.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -137.61. **CFL02.** Formula: $C_{11}H_{11}F_3N_2O_3$. M_w 276.2 Da. MS (ESI): $[M+H]^+_{calc}$ 277.1; $[M+H]^+_{obs}$ 277.1. HPLC: t_R 20.0 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -61.91. **CFL03.** Formula: $C_9H_6FNO_2$. M_w 179.15 Da. HPLC: 16.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -124.27. **CFL04.** Formula: $C_{12}H_{17}FN_2$. M_w 208.3 Da. HPLC: t_R 8.0 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -115.03. **~~CFL05~~.** Formula: $C_{13}H_{19}FN_2O$. M_w 238.3 Da. HPLC: t_R 8.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **CFL06. .** Formula: $C_{10}H_{13}FN_2O_2S$. M_w 244.3 Da. HPLC: t_R 12.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -109.79. **CFL07.** Formula: C_6H_5FO. M_w 112.1 Da. HPLC: t_R 13.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -137.61. **CFL08.** Formula: C_6H_4BrFO. M_w 191.0 Da. HPLC: t_R 16.9 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -123.63. **CFL09.** Formula: C_6H_4BrFO. M_w 191.0 Da. HPLC: t_R 17.2 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -114.13. **CFL10.** Formula: C_6H_4ClFO. M_w 146.5 Da. HPLC: t_R 16.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -114.61. **CFL11.** Formula: C_6H_5FO. M_w 112.1 Da. HPLC: t_R 14.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -112.33. **CFL12.** Formula: C_6H_5FO. M_w 112.1 Da. HPLC: t_R 14.0 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -124.90. **CFL13.** Formula: C_6H_4BrFO. M_w 191.0 Da.

HPLC: t_R 17.4 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -106.69. ~~**CFL14.**~~ Formula: C_6H_4ClFO. M_w 146.5 Da. HPLC: t_R 16.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **CFL15**. Formula: $C_8H_7FO_2$. M_w 154.1 Da. HPLC: t_R 18.2 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -105.18. ~~**CFL16.**~~ Formula: $C_7H_4F_4O$. M_w 180.1 Da. HPLC: t_R 17.9 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **CFL17.** Formula: $C_8H_5F_3N_2OS$. M_w 234.2 Da. HPLC: t_R 15.2 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -58.16. **CFL18.** Formula: $C_6H_5F_3N_2O$. M_w 178.1 Da. HPLC: t_R 11.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -70.78. **CFL19.** Formula: $C_5H_3F_3N_2OS$. M_w 196.1 Da. HPLC: t_R 12.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -71.10. **CFL20.** Formula: $C_7H_8F_3N_3O$. M_w 207.2 Da. HPLC: t_R 15.4 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -71.05. **CFL21.** Formula: $C_6H_5F_3N_2OS$. M_w 210.2 Da. HPLC: t_R 15.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -70.21. **CFL22.** Formula: $C_5H_4F_3N_3O$. M_w 179.1 Da. HPLC: t_R 17.5 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -71.20. **CFL23.** Formula: $C_5H_5F_3N_2O$. M_w 166.1 Da. HPLC: t_R 15.2 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -60.78. **CFL24.** Formula: $C_7H_7FO_2$. M_w 142.1 Da. HPLC: t_R 11.6 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -136.83.

Scored out compounds did not pass QC checks, therefore were not included in the final screening pools. **CFL Fragment Pools.** *CFL-Pool-1*: 01, 04, 11, 12, 15. *CFL-Pool-2*: 03, 06, 07, 10. *CFL-Pool-3*: 08, 09, 13, 24. *CFL-Pool-4*: 20, 21, 22, 23. *CFL-Pool-5*: 02, 17, 18, 19.

8.3.5. Directed Diversity Fragment Library (DDFL) QC Data

DDFL01. Formula: $C_{13}H_{15}F_3N_2O$. MS (ESI): $[M+H]^+_{calc}$ 273.1; $[M+H]^+_{obs}$ 273.0. HPLC t_R 13.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -68.81. **DDFL02.** Formula: $C_{11}H_7F_3N_2O_3$. MS (ESI): $[M+H]^+_{calc}$ 273.0; $[M+H]^+_{obs}$ 273.0. HPLC t_R 13.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -63.09. **DDFL03.** Formula: $C_8H_{10}F_3NO_3$. MS (ESI): $[M+H]^+_{calc}$ 226.1; $[M+H]^+_{obs}$ 226.0. HPLC t_R 13.9 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -68.77. **DDFL04.** Formula: $C_7H_4ClF_3N_2O_2$. MS (ESI): $[M+H]^+_{calc}$ 241.0; $[M+H]^+_{obs}$ 241.0. HPLC t_R 18.5 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -66.49. **DDFL05.** Formula: $C_9H_6ClF_3O_2$. MS (ESI): $[M+H]^+_{calc}$ 239.0; $[M+H]^+_{obs}$ N/A. HPLC t_R 18.6 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.40. **DDFL06.** Formula: $C_8H_9F_3N_2O_2$. MS (ESI): $[M+H]^+_{calc}$ 223.1; $[M+H]^+_{obs}$ N/A. HPLC t_R 17.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -68.80. **DDFL07.** Formula: $C_{11}H_{10}F_3N_3O$. MS (ESI): $[M+H]^+_{calc}$ 258.1; $[M+H]^+_{obs}$ 258.0. HPLC t_R 15.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -73.22. **DDFL08.** Formula: $C_{13}H_{18}F_3N_3$. MS (ESI): $[M+H]^+_{calc}$ 274.2; $[M+H]^+_{obs}$ 274.0. HPLC t_R 10.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -58.23. **DDFL09.** Formula: $C_8H_6F_3NO_2$. MS (ESI): $[M+H]^+_{calc}$ 206.0; $[M+H]^+_{obs}$ 206.0. HPLC t_R 15.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.86. **DDFL10.** Formula: $C_8H_5F_3N_4$. MS (ESI): $[M+H]^+_{calc}$ 215.0; $[M+H]^+_{obs}$ 215.0. HPLC t_R 16.0 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -61.99. ~~**DDFL11.**~~ Formula: $C_9H_6F_3N_3O$. MS (ESI): $[M+H]^+_{calc}$ 230.0; $[M+H]^+_{obs}$ N/A. HPLC t_R 12.0 mins; > 80 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -74.79, -75.39. **DDFL12.** Formula: $C_8H_7F_3N_2O$. MS (ESI): $[M+H]^+_{calc}$ 205.1; $[M+H]^+_{obs}$ 205.0. HPLC t_R 8.3 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -58.73. **DDFL13.** Formula: $C_8H_7F_3N_2O$. MS (ESI): $[M+H]^+_{calc}$ 205.1; $[M+H]^+_{obs}$ 205.1. HPLC t_R 13.4 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -57.58. **DDFL14.** Formula: $C_3H_5F_3O_3S$. MS (ESI): $[M+H]^+_{calc}$ 179.0; $[M+H]^+_{obs}$ 178.7. HPLC t_R N/A. NMR: ^{19}F (376

MHz, 10 % D_2O) δ ppm -78.66. **DDFL15.** Formula: $C_6H_9F_3N_2O$. MS (ESI): $[M+H]^+_{calc}$ 183.1; $[M+H]^+_{obs}$ 183.1. HPLC t_R 4.4 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -75.51. **DDFL16.** Formula: $C_8H_{11}F_3N_2$. MS (ESI): $[M+H]^+_{calc}$ 193.1; $[M+H]^+_{obs}$ 193.0. HPLC t_R 19.9 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -61.35. **DDFL17.** Formula: $C_8H_9F_3N_2$. MS (ESI): $[M+H]^+_{calc}$ 191.1; $[M+H]^+_{obs}$ N/A. HPLC t_R 13.5 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.14. **DDFL18.** Formula: $C_6H_3F_3N_2O_2$. MS (ESI): $[M+H]^+_{calc}$ 193.0; $[M+H]^+_{obs}$ N/A. HPLC t_R 14.0 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -70.03. **DDFL19.** Formula: $C_8H_8F_3N_3O_2$. MS (ESI): $[M+H]^+_{calc}$ 236.1; $[M+H]^+_{obs}$ 236.1. HPLC t_R 12.3 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -76.62. **DDFL20.** Formula: $C_6H_5F_3N_2O$. MS (ESI): $[M+H]^+_{calc}$ 179.0; $[M+H]^+_{obs}$ 178.7. HPLC t_R 11.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -70.78. **DDFL21.** Formula: $C_{12}H_{13}F_3N_2O_2$. MS (ESI): $[M+H]^+_{calc}$ 275.1; $[M+H]^+_{obs}$ 275.1. HPLC t_R 14.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -60.99. **DDFL22.** Formula: $C_9H_9ClF_3N_3$. MS (ESI): $[M+H]^+_{calc}$ 252.0; $[M+H]^+_{obs}$ 252.1. HPLC t_R 23.2 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -70.20. **DDFL23.** Formula: $C_8H_8F_3N_3$. MS (ESI): $[M+H]^+_{calc}$ 204.1; $[M+H]^+_{obs}$ 204.1. HPLC t_R 13.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.13. **DDFL24.** Formula: $C_7H_9F_3N_4O$. MS (ESI): $[M+H]^+_{calc}$ 223.1; $[M+H]^+_{obs}$ 223.0. HPLC N/A. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -60.21. ~~**DDFL25.**~~ Formula: $C_7H_8F_3NO_2$. MS (ESI): $[M+H]^+_{calc}$ 196.1; $[M+H]^+_{obs}$ N/A. HPLC t_R 11.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -68.80, -69.30. **DDFL26.** Formula: $C_{11}H_7F_3N_2O_3$. MS (ESI): $[M+H]^+_{calc}$ 273.0; $[M+H]^+_{obs}$ 273.0. HPLC t_R 16.9 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.98. ~~**DDFL27.**~~ Formula: $C_6H_9F_3N_2O$. MS (ESI): $[M+H]^+_{calc}$ 182.1; $[M+H]^+_{obs}$ N/A. HPLC t_R 4.6 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -69.00, -75.40. **DDFL28.** Formula: $C_9H_9F_3O_2S$. MS (ESI): $[M+H]^+_{calc}$ 239.0; $[M+H]^+_{obs}$ N/A. HPLC t_R 17.5 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -57.61. **DDFL29.** Formula: $C_{11}H_{12}F_3N_3O$. MS (ESI): $[M+H]^+_{calc}$ 260.1; $[M+H]^+_{obs}$ 260.0. HPLC t_R 15.4 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -61.08. **DDFL30.** Formula: $C_{11}H_9F_3N_2O$. MS (ESI): $[M+H]^+_{calc}$ 243.1; $[M+H]^+_{obs}$ 243.0. HPLC t_R 13.4 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.34. **DDFL31.** Formula: $C_9H_7F_3O_2$. MS (ESI): $[M+H]^+_{calc}$ 205.0; $[M+H]^+_{obs}$ N/A. HPLC t_R 19.5 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -57.46. **DDFL32.** Formula: $C_{11}H_{13}F_3N_4O$. MS (ESI): $[M+H]^+_{calc}$ 275.1; $[M+H]^+_{obs}$ 275.1. HPLC t_R 16.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -70.44. **DDFL33.** Formula: $C_8H_{11}F_3O_3$. MS (ESI): $[M+H]^+_{calc}$ 213.1; $[M+H]^+_{obs}$ N/A. HPLC N/A. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -84.19. **DDFL34.** Formula: $C_9H_{10}F_3N$. MS (ESI): $[M+H]^+_{calc}$ 190.1; $[M+H]^+_{obs}$ 190.0. HPLC t_R 12.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -77.10. **DDFL35.** Formula: $C_{10}H_{13}F_3N_4$. MS (ESI): $[M+H]^+_{calc}$ 247.1; $[M+H]^+_{obs}$ 247.1. HPLC t_R 14.2 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -70.42. **DDFL36.** Formula: $C_9H_8F_3N_3O_2$. MS (ESI): $[M+H]^+_{calc}$ 248.1; $[M+H]^+_{obs}$ N/A. HPLC N/A. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.43. ~~**DDFL37.**~~ Formula: $C_5H_5F_3N_2$. MS (ESI): $[M+H]^+_{calc}$ 151.0; $[M+H]^+_{obs}$ 151.0. HPLC t_R 15.6 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **DDFL38.** Formula: $C_{11}H_9F_3N_2O$. MS (ESI): $[M+H]^+_{calc}$ 243.1; $[M+H]^+_{obs}$ 243.1. HPLC t_R 18.0 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.31. **DDFL39.** Formula: $C_5H_4F_3N_3O$. MS (ESI): $[M+H]^+_{calc}$ 180.0; $[M+H]^+_{obs}$ N/A. HPLC t_R 11.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -71.19. **DDFL40.** Formula: $C_{10}H_7F_3N_2O_2S$. MS (ESI): $[M+H]^+_{calc}$ 278.0; $[M+H]^+_{obs}$ 277.0. HPLC t_R 17.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -58.67. ~~**DDFL41.**~~ Formula: $C_{11}H_9F_3N_2O_2S$. HPLC t_R 20.9 mins; > 75 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **DDFL42.** Formula: $C_8H_5F_3N_2$. HPLC t_R 17.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -68.20. **DDFL43.** Formula: $C_{10}H_8F_3N_5$. HPLC t_R 14.9 mins; > 95 % purity.

NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.37. **~~DDFL44.~~** Formula: $C_4H_4F_3N_3$. HPLC t_R 10.7 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **DDFL45.** Formula: $C_6H_2F_3Cl_2N$. HPLC t_R 19.8 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -56.56. **DDFL46.** Formula: $C_9H_9F_3N_2O$. HPLC t_R 13.4 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.76. **~~DDFL47.~~** Formula: $C_9H_5F_3N_2O$. HPLC t_R 19.7 mins; > 85 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -65.50, -73.18. **DDFL48.** Formula: $C_{12}H_{12}F_3NO_3$. HPLC t_R 17.6 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.32. **~~DDFL49~~.** Formula: $C_5H_2F_3N_5O_2$. HPLC t_R 17.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -65.50, -75.40. **DDFL50.** Formula: $C_8H_6ClF_3O$. HPLC t_R 22.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.59. **DDFL51.** Formula: $C_{11}H_{12}F_3N_3$. HPLC t_R 19.4 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -57.69. **DDFL52.** Formula: $C_8H_5F_3O_2$. HPLC t_R 16.5 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.70. **DDFL53.** Formula: $C_{11}H_6F_3NO_3$. HPLC t_R 16.6 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -61.81. **DDFL54.** Formula: $C_4H_5F_3O$. HPLC N/A. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -67.33. **DDFL55.** Formula: $C_{12}H_{14}F_3NO$. HPLC t_R 14.1 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.17. **DDFL56.** Formula: $C_8H_{11}F_3N_2$. HPLC t_R 19.0 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -61.73. **DDFL57.** Formula: $C_9H_5F_3N_2O_3$. HPLC t_R 16.2 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -57.96. **DDFL58.** Formula: $C_{12}H_{10}F_3N_3O_2$. HPLC t_R 18.0 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -66.32. **DDFL59.** Formula: $C_7H_6F_3NO$. HPLC t_R 8.5 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -62.47. **~~DDFL60~~.** Formula: $C_{12}H_{16}FN$. HPLC t_R 12.2 mins; > 95 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **~~DDFL61.~~** Formula: C_6H_6FN. HPLC N/A. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **~~DDFL62.~~** Formula: $C_6H_{11}FO_2$. HPLC N/A. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm N/A. **~~DDFL63.~~** Formula: C_7H_6FN. HPLC t_R 13.4 mins; < 50 % purity. NMR: ^{19}F (376 MHz, 10 % D_2O) δ ppm -75.40, -119.66.

Scored out compounds did not pass QC checks, therefore were not included in the final screening pools. **DDFL Fragment Pools.** *DDFL-Pool-1:* 07, 10, 21, 22, 26, 33, 42, 45, 48, 57, 59. *DDFL-Pool-2:* 09, 14, 18, 24, 36, 38, 39, 51, 53, 54. *DDFL-Pool-3:* 01, 04, 05, 12, 20, 28, 34, 46, 55, 56. *DDFL-Pool-4*: 06, 13, 16, 17, 19, 32, 40, 43, 52, 58. *DDFL-Pool-5:* 02, 03, 8, 15, 23, 29, 30 ,31, 35, 50.

8.3.6. ^{19}F NMR Fragment Screening of UbE2L3

8.3.6.1. RAFL Fragment Screening of UbE2L3

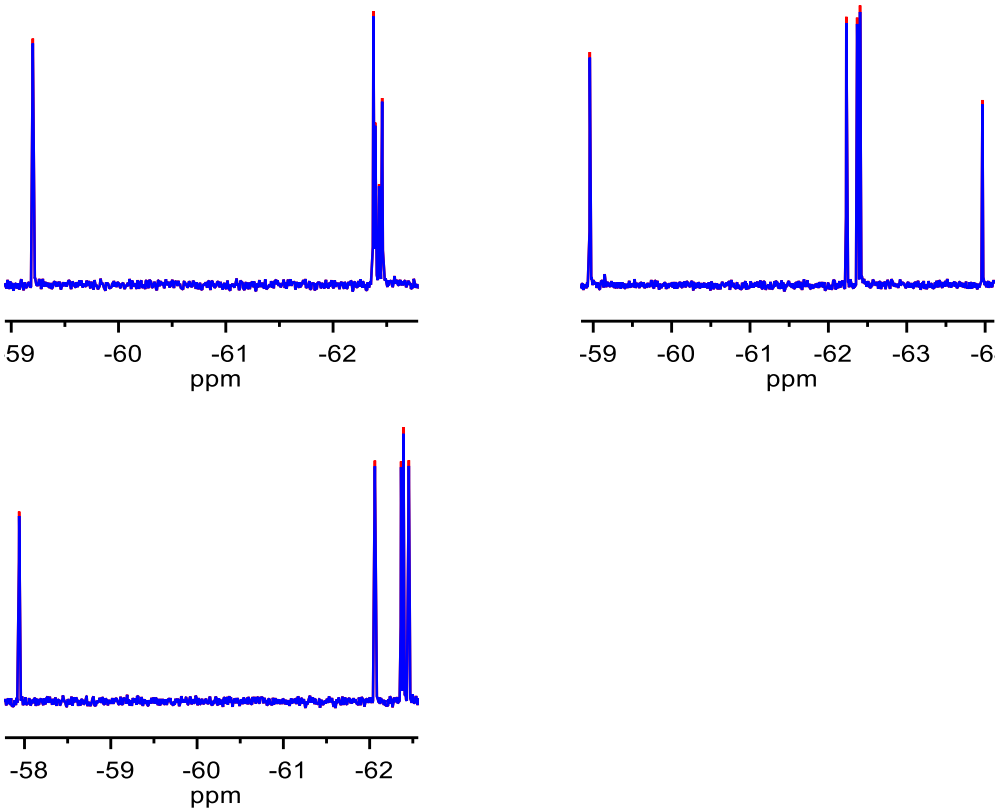


Figure 8.21: 1D ^{19}F NMR spectra of RAFL Screening against UbE2L3. Superposition of fragments in absence (red) and presence (blue) of 10 μM UbE2L3 protein. Top left: RAFL-Pool-1. Top right: RAFL-Pool-2. Bottom left: RAFL-Pool-3.

8.3.6.2. CFL Fragment Screening of UbE2L3

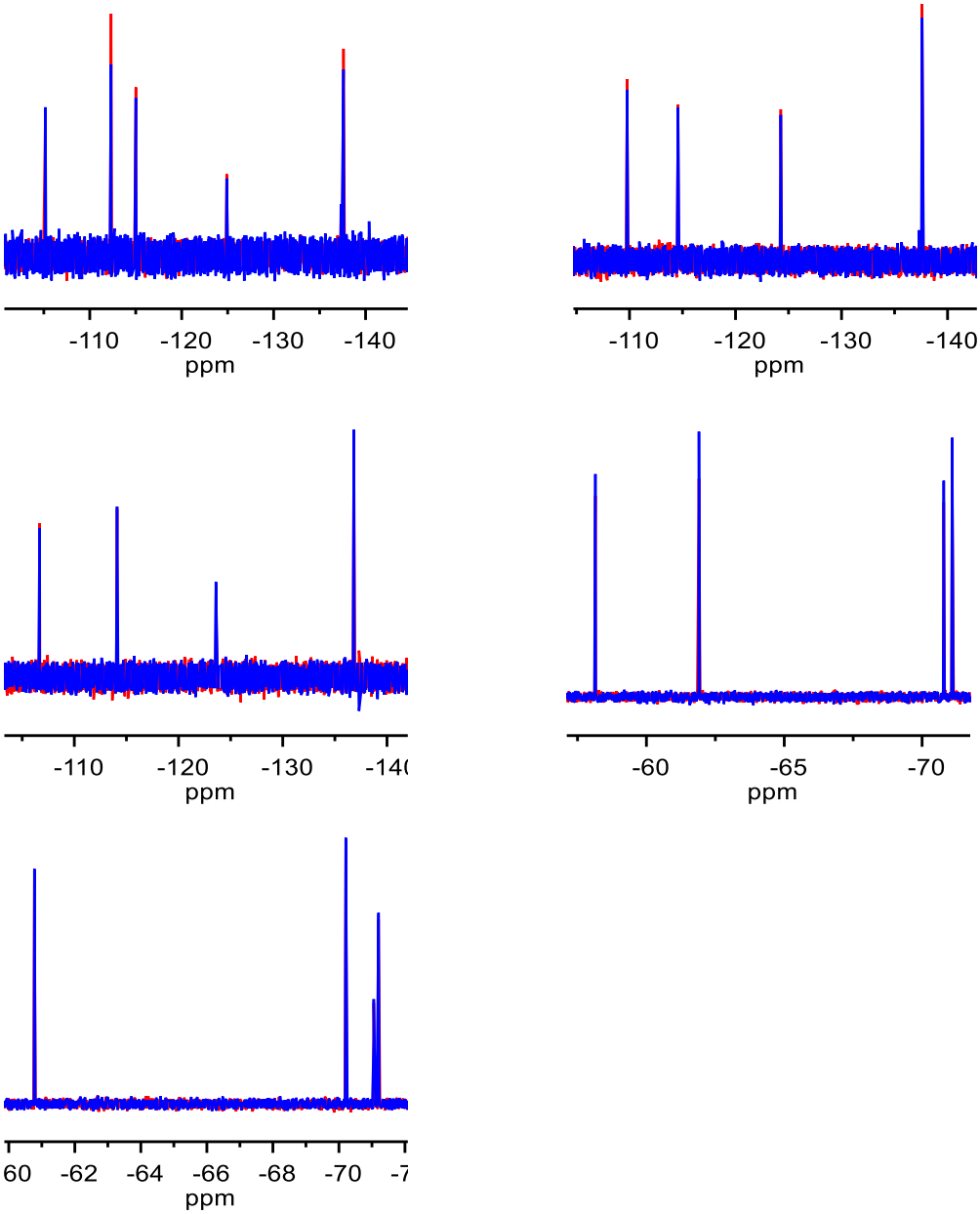


Figure 8.22: 1D ^{19}F NMR spectra of CFL screening against UbE2L3. Superposition of fragments in absence (red) and presence (blue) of 10 μM UbE2L3 protein. Top Left: CFL-Pool-1. Top right: CFL-Pool-2. Middle left: CFL-Pool-3. Middle right: CFL-Pool-4. Bottom left: CFL-Pool-5.

8.3.6.3. DDFL Fragment Screening of UbE2L3

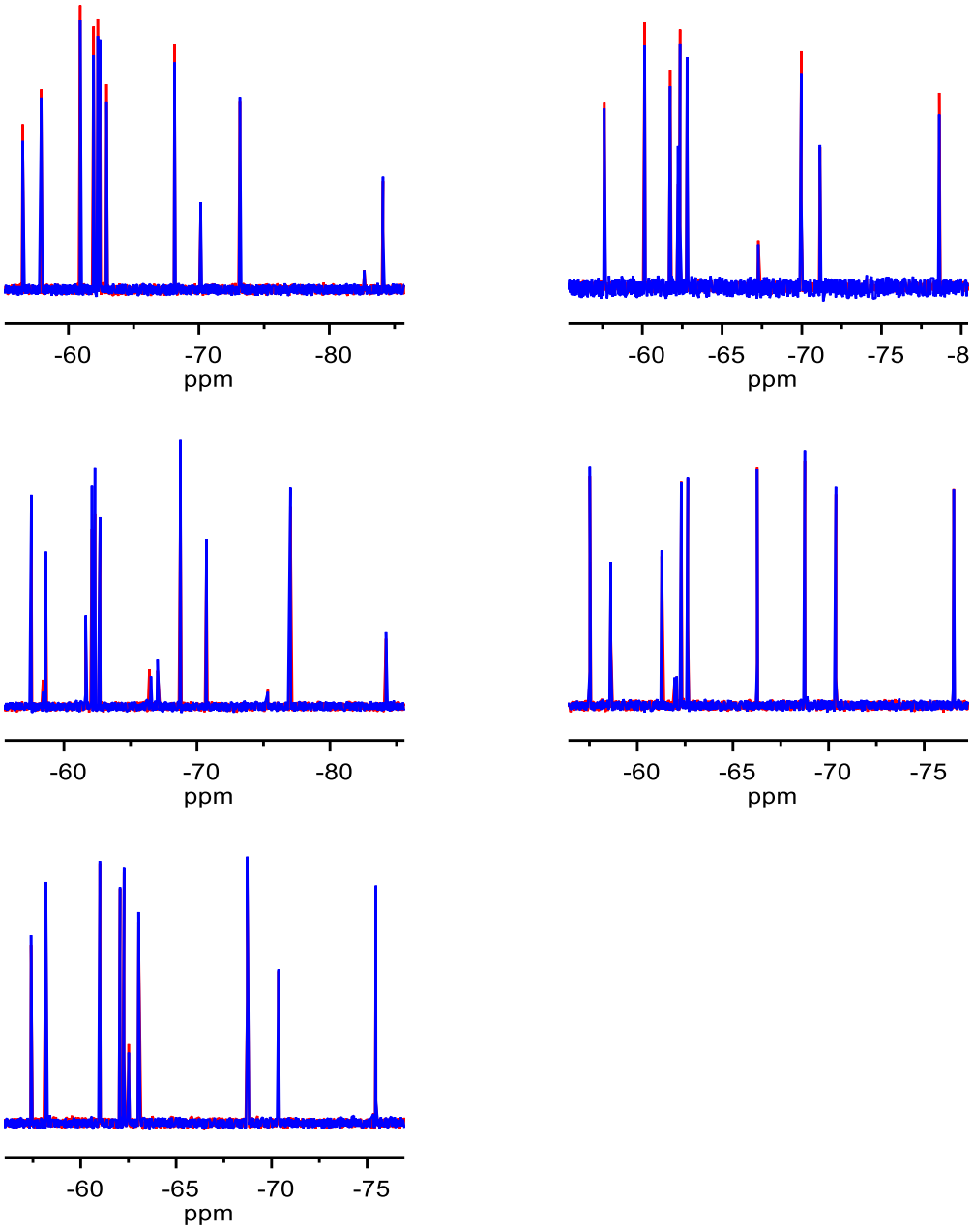


Figure 8.23: 1D ^{19}F NMR spectrum of DDFL Pool 1. Superposition of fragments in absence (red) and presence (blue) of 10 μM UbE2L3 protein. Top Left: DDFL-Pool-1. Top right: DDFL-Pool-2. Middle left: DDFL-Pool-3. Middle right: DDFL-Pool-4. Bottom left: DDFL-Pool-5.

8.4. Appendix for chapter 5

8.4.1. ^{19}F NMR Fragment Screening of NusEΔ/NusB

8.4.1.1. RAFL Fragment Screening of NusEΔ/NusB

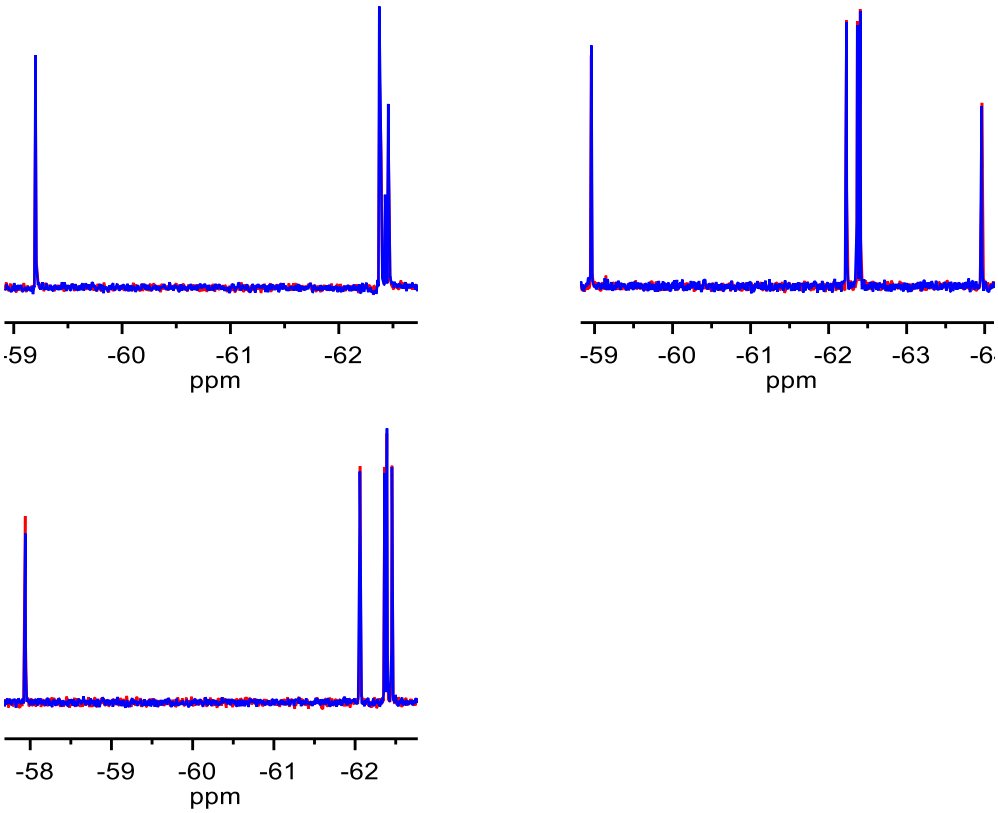
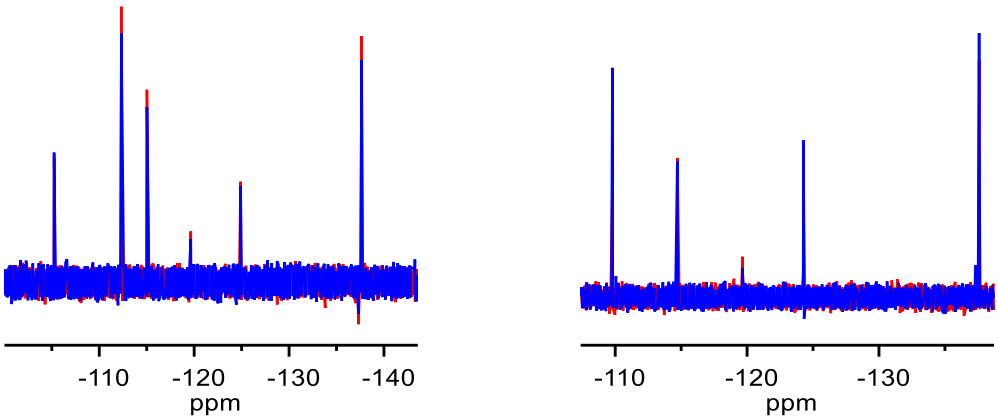


Figure 8.24: 1D ^{19}F NMR spectra of RAFL Screening against NusEΔ/NusB. Superposition of fragments in absence (red) and presence (blue) of 10 μM protein. Top left: RAFL-Pool-1. Top right: RAFL-Pool-2. Bottom left: RAFL-Pool-3.

8.4.1.2. CFL Fragment Screening of NusEΔ/NusB



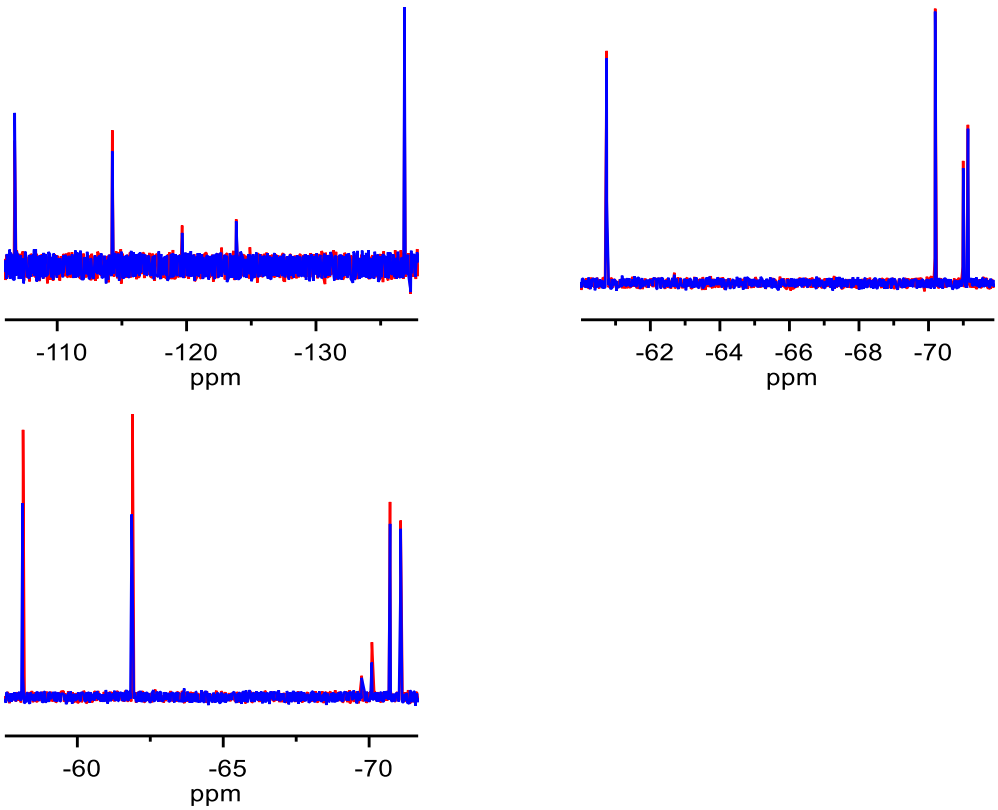


Figure 8.25: 1D ^{19}F NMR spectra of CFL Screening against NusEΔ/NusB. Superposition of fragments in absence (red) and presence (blue) of 10 μM NusEΔ/NusB protein. Top Left: CFL-Pool-1. Top right: CFL-Pool-2. Middle left: CFL-Pool-3. Middle right: CFL-Pool-4. Bottom left: CFL-Pool-5.

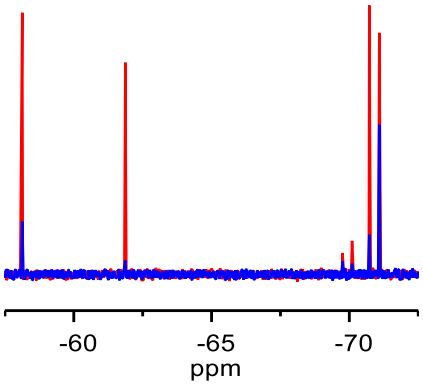


Figure 8.26: 1D ^{19}F CPMG NMR spectra of CFL-Pool-5. Left: compound only spectra. Right: containing 5 μM NusEΔ/NusB. (Red spectra: T_{CPMG} = 20 ms. Blue spectra: T_{CPMG} = 400 ms).

8.4.1.3. DDFL Fragment Library Screening of NusEΔ/NusB

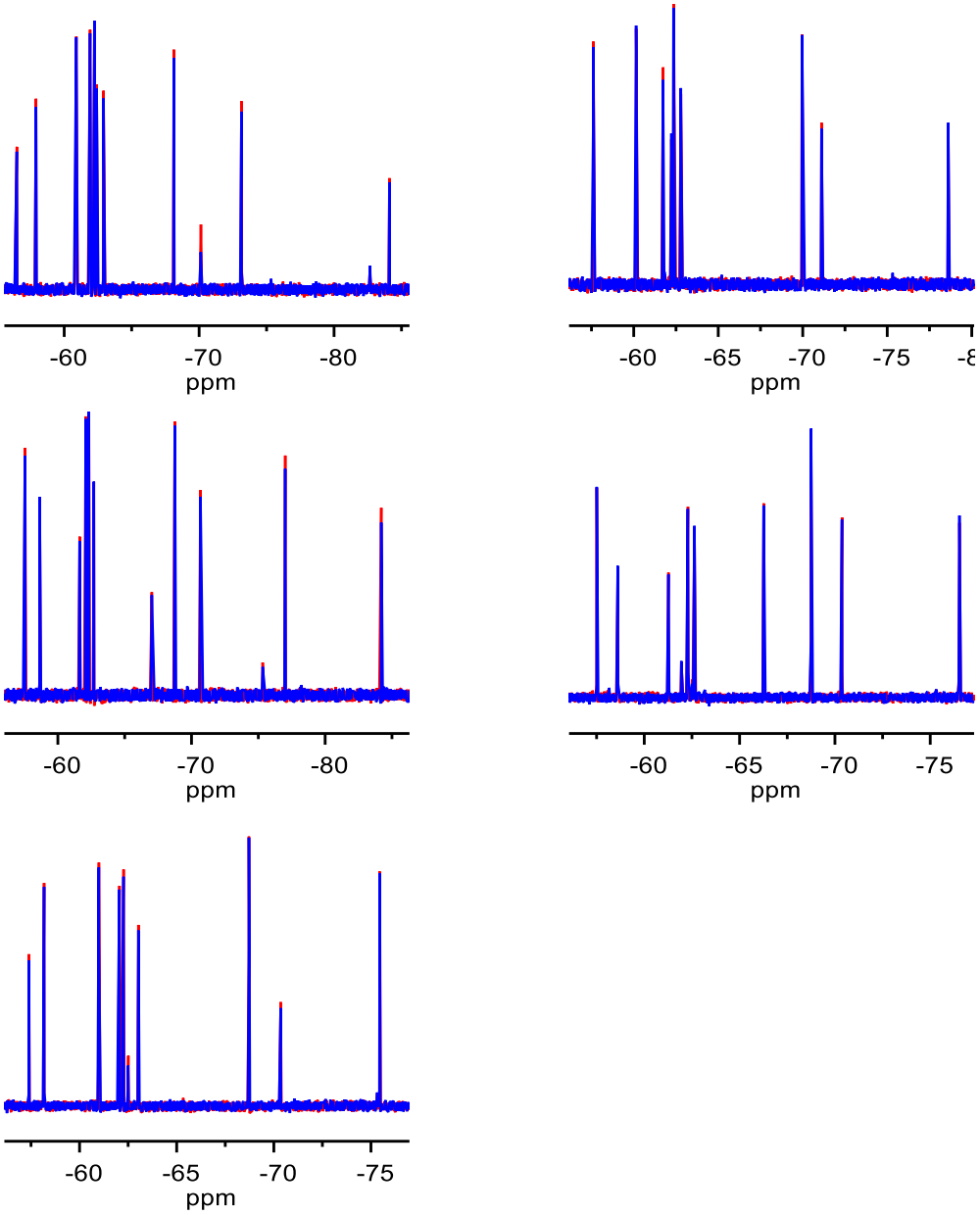
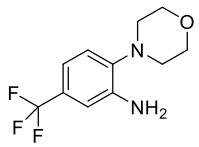
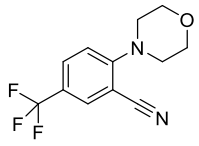


Figure 8.27: 1D ^{19}F NMR spectrum of DDFL Pool 1. Superposition of fragments in absence (red) and presence (blue) of 10 μM NusEΔ/NusB protein. Top Left: DDFL-Pool-1. Top right: DDFL-Pool-2. Middle left: DDFL-Pool-3. Middle right: DDFL-Pool-4. Bottom left: DDFL-Pool-5.

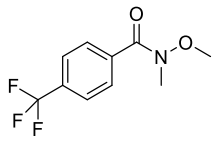
8.4.2. Compound Analogue structures and QC data



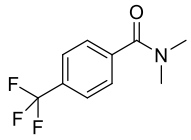
CFL02a. Formula: $C_{11}H_{13}F_3N_2O$. HPLC t_R 18.0 mins, > 95 % purity. NMR: ^{1}H (500 MHz, DMSO-d_6) δ ppm 7.01 (d, 1H, Ar**H**), 6.98 (d, 1H, Ar**H**), 6.85 (m, 1H, Ar**H**), 5.19 (s, 2H, N**H**$_2$), 3.77 (m, 4H, C**H**$_2$), 2.84 (m, 4H, C**H**$_2$). ^{13}C (125 MHz, DMSO-d_6) δ ppm 143.27 (s, 1C), 141.48 (s, 1C), 125.25 (q, 1C), 124.81(q, 1C), 119.59 (s, 1C), 113.35 (q, 1C), 110.68 (q, 1C), 66.88 (s, 2C), 50.72 (s, 2C). ^{19}F NMR (376 MHz, 10 % D_2O) δ ppm -61.78.



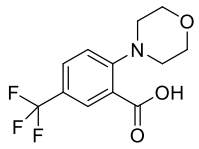
CFL02b. Formula: $C_{12}H_{11}F_3N_2O$. HPLC t_R 19.5 mins, > 95 % purity. NMR: ^{1}H NMR (500 MHz, DMSO) δ 8.13 (d, 1H, Ar**H**), 7.91 (dd, 1H, Ar**H**), 7.31 (d, 1H, Ar), 3.78 (m, 4H, C**H**$_2$), 3.32 (m, 4H, C**H**$_2$). ^{13}C NMR (126 MHz, DMSO-d_6) δ ppm 157.76 (s, 1C), 132.38 (q, 1C), 131.45 (q, 1C), 124.01 (q, 1C), 121.88 (q, 1C), 119.79 (s, 1C), 117.74 (s, 1C), 103.77 (s, 1C), 66.40 (s, 4C), 51.16 (s, 4C). ^{19}F NMR (376 MHz, 10 % D_2O) δ ppm -62.07.



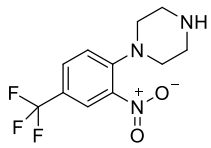
CFL02c. Formula: $C_{10}H_{10}F_3NO_2$. HPLC t_R 17.7 mins, > 95 % purity. NMR: ^{1}H NMR (500 MHz, DMSO-d_6) δ ppm 7.81 (q, 4H, Ar**H**), 3.33 (s, 3H, C**H**$_3$), 3.29 (s, 3H, C**H**$_3$). ^{13}C NMR (126 MHz, DMSO-d_6) δ ppm 139.02 (C), 130.74 (C), 128.83 (C), 125.54 (C), 61.41 (C). ^{19}F NMR (376 MHz, 10 % D_2O) δ ppm -62.67.



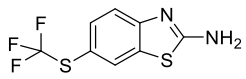
CFL02d. Formula: $C_{10}H_{10}F_3NO$. HPLC t_R 16.9 mins, > 95 % purity. NMR: ^{1}H NMR (500 MHz, DMSO-d_6) δ ppm 7.72 (dd, 4H, Ar**H**), 3.01 (s, 3H, C**H**$_3$), 2.88 (s, 3H, C**H**$_3$). ^{13}C NMR (126 MHz, DMSO-d_6) δ ppm 169.23 (s, 1C), 141.10 (s, 1C), 129.95 (q, 1C), 128.18 (s, 2C), 125.80 (q, 2C), 124.40 (q, 1C), 39.27 (s, 1C), 35.14 (s, 1C). ^{19}F NMR (376 MHz, 10 % D_2O) δ ppm -62.54.



CFL02e. Formula: $C_{12}H_{12}F_3NO_3$. HPLC t_R 15.6 mins, > 95 % purity. NMR: ^{1}H NMR (500 MHz, DMSO-d_6) δ ppm 13.89 (b. s, 1C, COO**H**), 7.94 (d, 1H, Ar**H**), 7.80 (dd, 1H, Ar**H**), 7.36 (d, 1H, Ar**H**), 3.74 (t, 4H, C**H**$_2$), 3.12 (t, 4H, C**H**$_2$). ^{13}C NMR (126 MHz, DMSO-d_6) δ ppm 167.95 (s, 1C), 154.18 (s, 1C), 129.67 (q, 1C), 128.34 (q, 1C), 124.52 (q, 1C), 124.33 (s, 1C), 121.85 (q, 1C), 120.31 (s, 1C), 66.45 (s, 1C), 52.01 (s, 1C). ^{19}F NMR (376 MHz, 10 % D_2O) δ -61.30.



CFL02f. Formula: $C_{11}H_{12}F_3N_3O_2$. MS (ESI): $[M+H]^+_{calc}$ 276.1; $[M+H]^+_{obs}$ 276.1. HPLC t_R 15.3 mins, > 95 % purity. NMR: ^{1}H NMR (500 MHz, DMSO-d_6) δ ppm 8.12 (d, 1H, Ar**H**), 7.82 (dd, 1H, Ar**H**), 7.41 (d, 1H, Ar**H**), 3.05 (t, 4H, C**H**$_2$), 2.80 (t, 4H, C**H**$_2$). ^{13}C NMR (126 MHz, DMSO-d_6) δ ppm 148.34 (s, 1C), 139.82 (s, 1C), 130.51 (q, 1C), 124.20 (q, 1C), 124.17 (q, 1C), 121.73 (s, 1C), 119.48 (q, 1C), 51.95 (s, 2C), 45.70 (s, 2C). ^{19}F NMR (376 MHz, 10 % D_2O) δ ppm -62.06.



CFL17a. Formula: $C_8H_5F_3N_2S_2$. MS (ESI): $[M+H]^+_{calc}$ 251.0; $[M+H]^+_{obs}$ 251.1. HPLC t_R 16.4 mins, > 95 % purity. NMR: ^{1}H NMR (500 MHz, DMSO-d_6) δ ppm 8.08 (d, 1H, Ar**H**), 7.86 (s, 2H, N**H**$_2$), 7.51 (dd, 1H, Ar**H**), 7.41 (d, 1H, Ar**H**). ^{19}F NMR (376 MHz, 10 % D_2O) δ ppm -43.87.

8.4.3. Microdialysis (Equation derived by Dr Steven Shave)

$$K_D = \frac{-[L]_0 r_b^2 V_1 - [L]_0 r_b^2 V_2 + [L]_0 V_1 r_b r_o + [L]_0 V_2 r_b r_o - [P]_0 V_1 r_b r_o - [P]_0 V_2 r_b}{(r_b - r_o)(r_o V_1 + V_2)}$$

Variable	Description
K_D	Dissociation constant
$[L]_0$	Total ligand concentration over whole system
$[P]_0$	Total protein concentration in red well
V_1	Red volume
V_2	White volume
r_b	Ratio of ligand in red/white chamber in absence of protein $\left(\frac{[L]_{red}}{[L]_{white}}\right)_{no\ protein}$
r_o	Ratio of ligand in red/white chamber in presence of protein $\left(\frac{[L]_{red}}{[L]_{white}}\right)_{with\ protein}$

Raw Data (CFL02f)

Control experiment				
Well Number	**Peak intensity**	**pmoles**	**Concentration (μM)**	**Intensity Ratio**
Red 1	5.01 x10⁶	0.815	0.408	1.008
White 1	4.97 x10⁶	0.808	0.404	
Red 2	6.24 x10⁶	1.020	0.508	0.943
White 2	6.62 x10⁶	1.080	0.538	
Red 3	6.95 x10⁶	1.130	0.565	1.012
White 3	6.87 x10⁶	1.120	0.559	
			Average	0.987
With 40 μM total protein concentration				
Well Number	**Peak intensity**	**pmoles**	**Concentration (μM)**	**Intensity Ratio**
Red 1	5.17 x10⁶	0.841	0.421	1.222
White 1	4.23 x10⁶	0.688	0.344	
Red 2	6.79 x10⁶	1.100	0.552	1.157
White 2	5.87 x10⁶	0.955	0.477	
Red 3	9.83 x10⁶	1.600	0.800	1.255
White 3	7.83 x10⁶	1.270	0.637	
			Average	1.211

Table 8.4: Raw data from LCMS analysis of qμD for compound CFL02f.

8.4.4. Bacterial inhibition assay

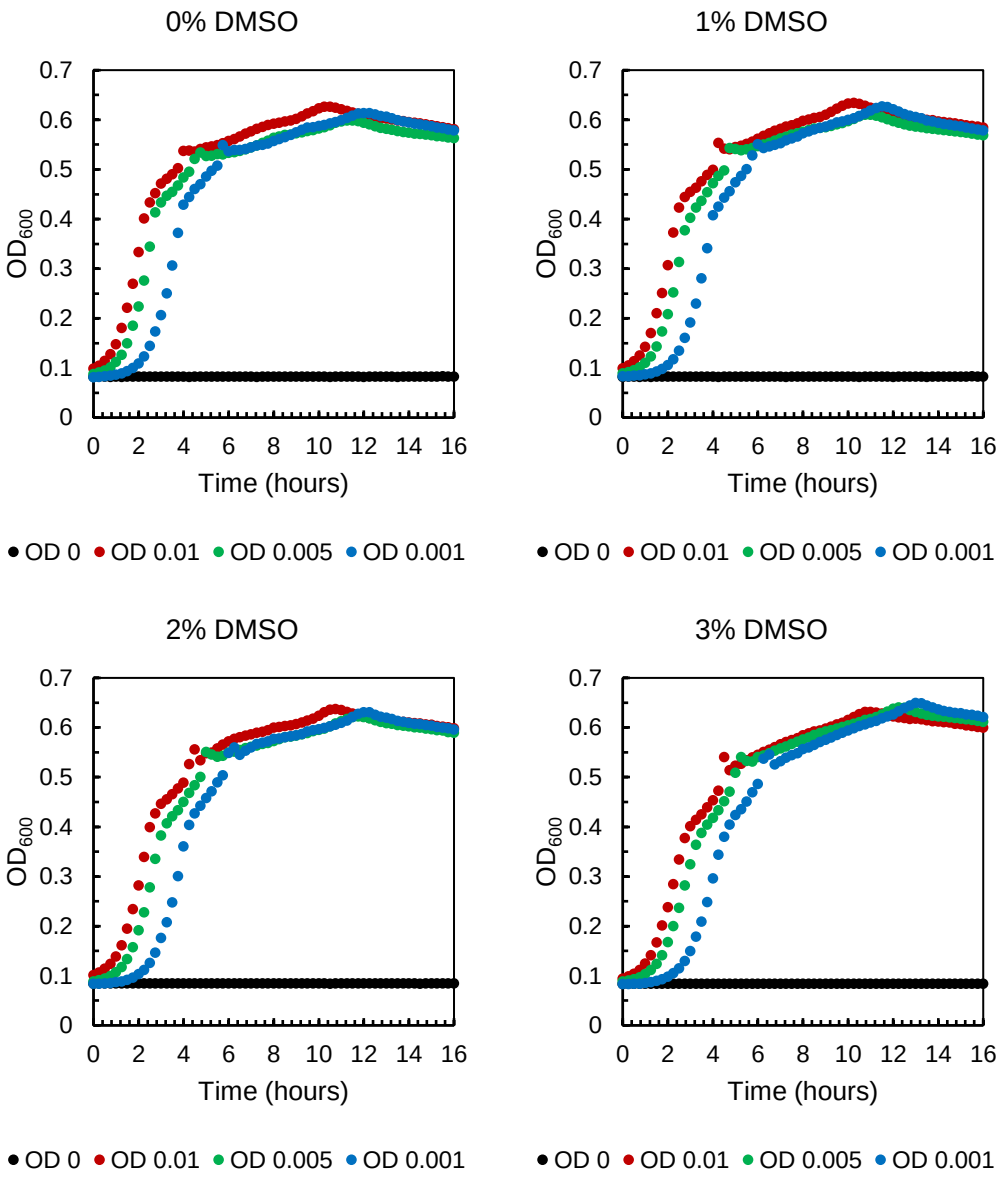
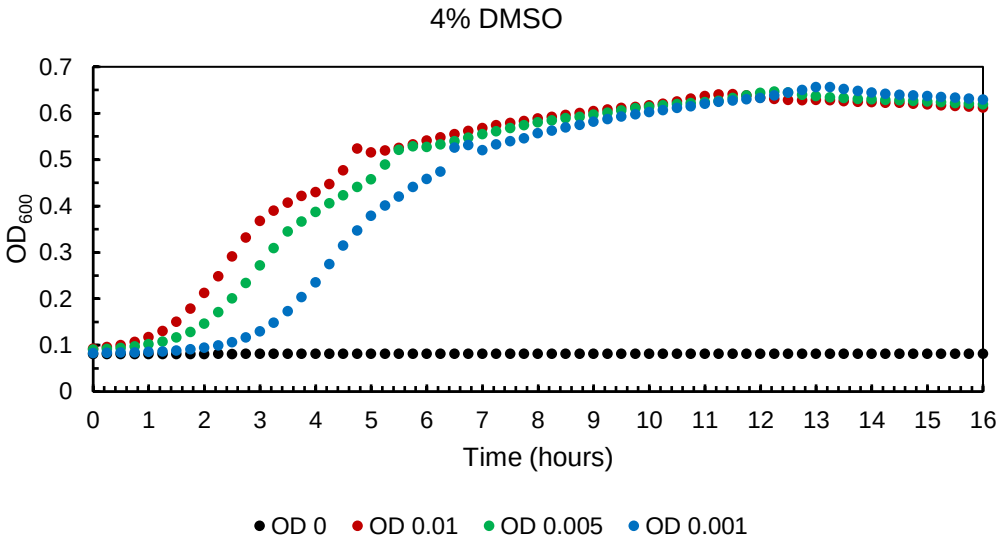


Figure 8.28: Plots of bacterial growth curves under different conditions. The starting OD_{600} is varied from 0, 0.001, 0.005, 0.01. The DMSO concentration range is from 0, 1, 2, 3, 4 %.



8.4.5. ^{15}N HSQCs from all titrations with active CFL compounds

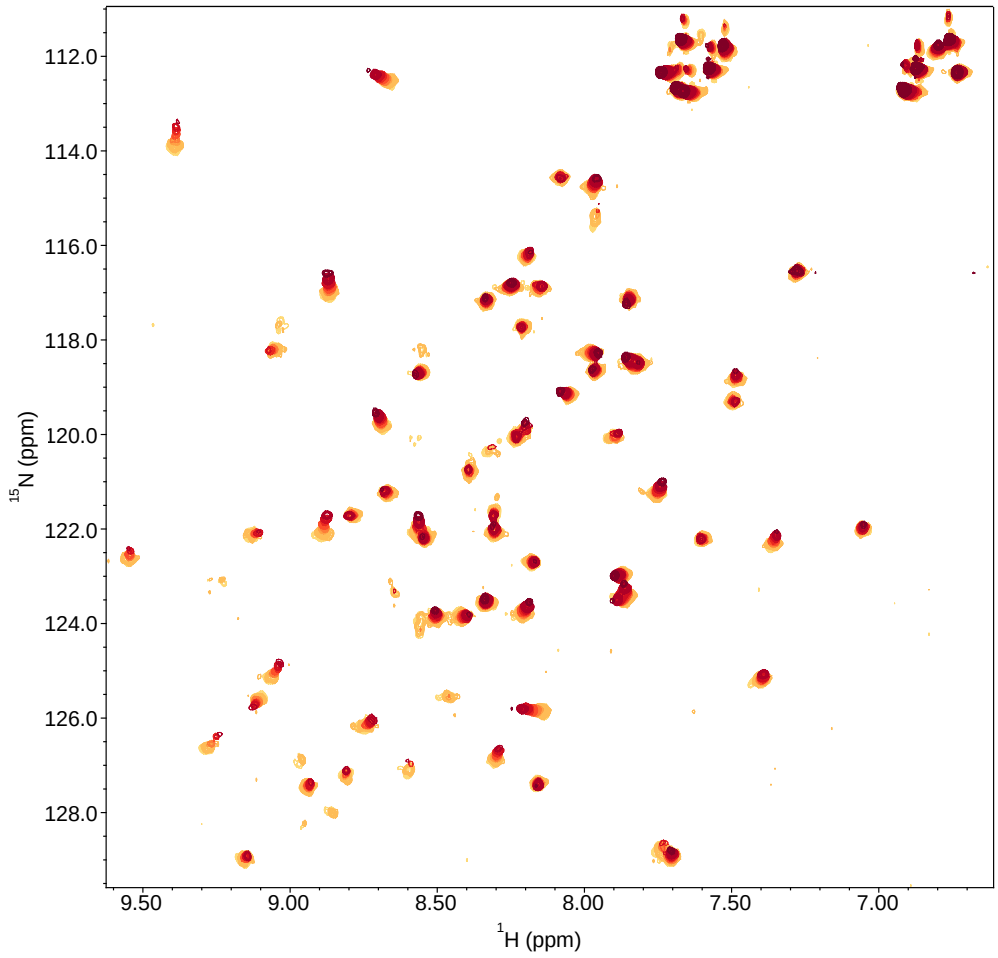


Figure 8.29: [^{1}H, ^{15}N] HSQC DMSO titration of ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex with DMSO (0 – 10 %; light to dark).

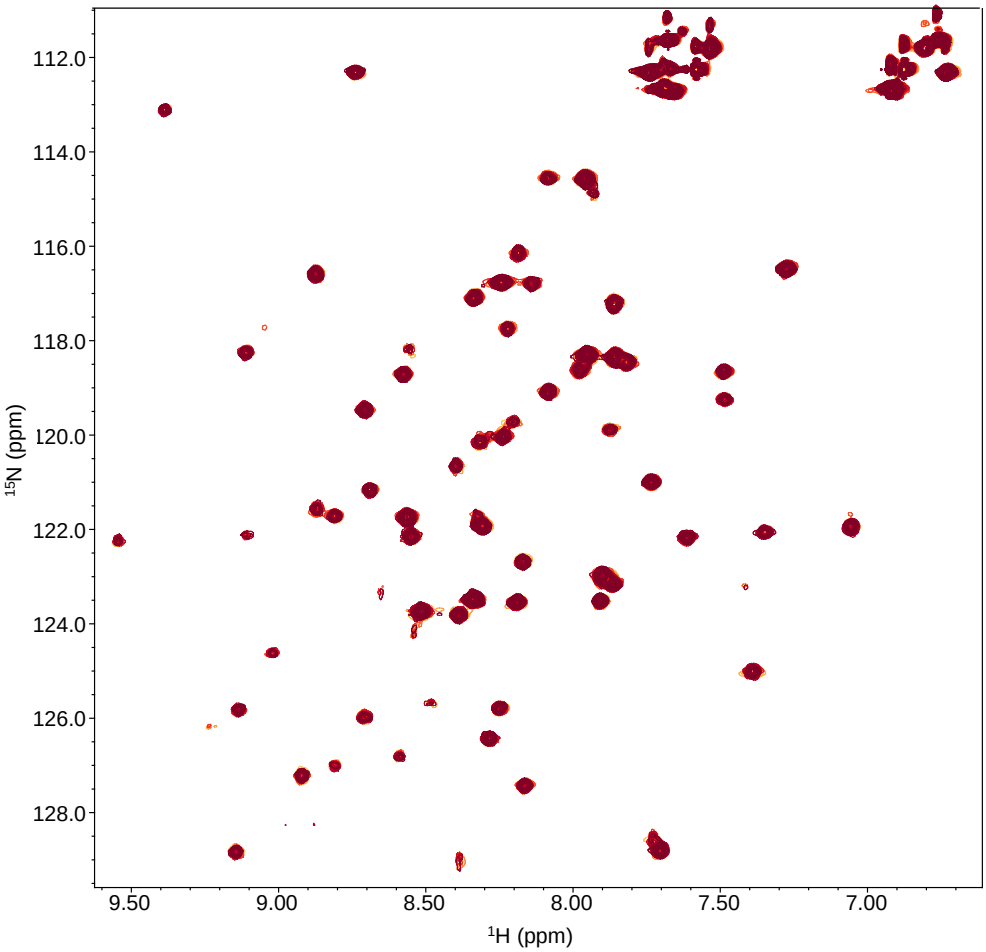


Figure 8.30: [^{1}H, ^{15}N] HSQC DMSO stability test of ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex. NusEΔ/NusB (100 μM) in the presence of 10 % DMSO. 13 spectra acquired over 72 hours (t(0) = light to dark).

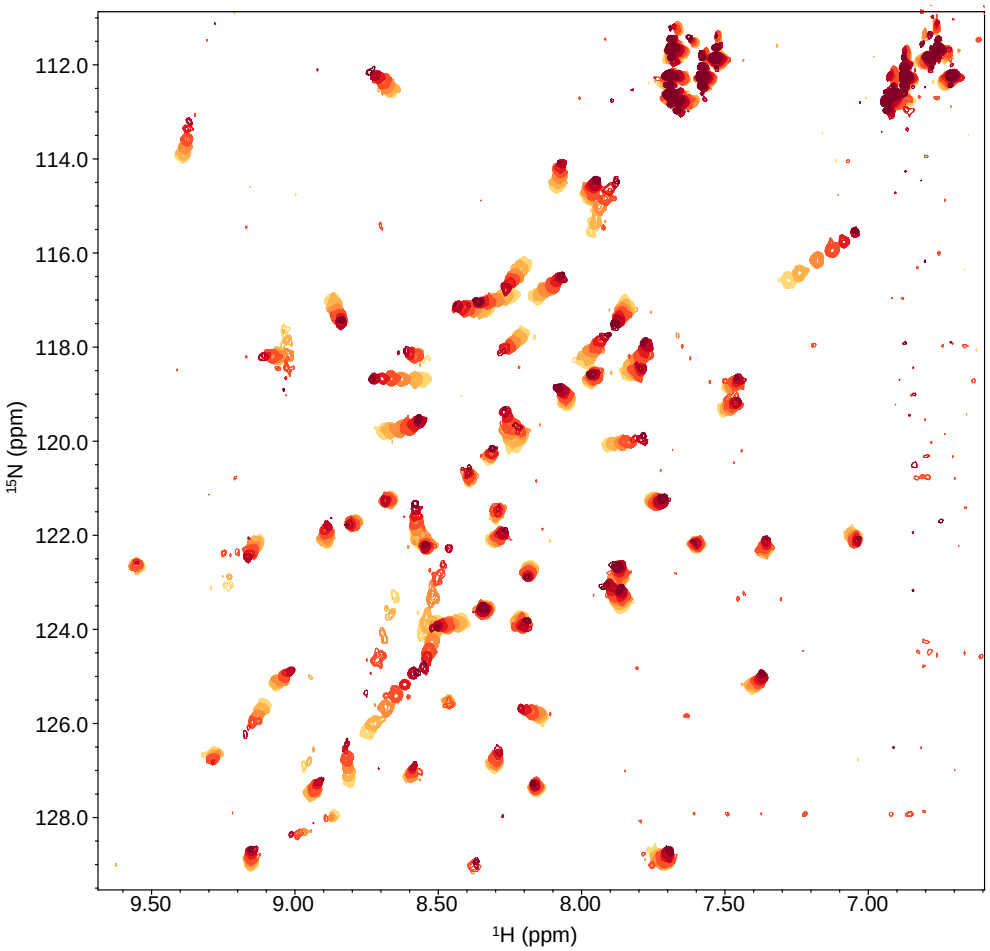


Figure 8.31: [^{1}H, ^{15}N] HSQC spectrum of a series of titration points with compound CFL02 and ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex. From light to dark shows the change in chemical shift on increasing concentration of ligand. Concentration of ligand ([L]0) ranges from 0 μM to 1.9 mM. 7 titration points were collected in total and the CSPs monitored.

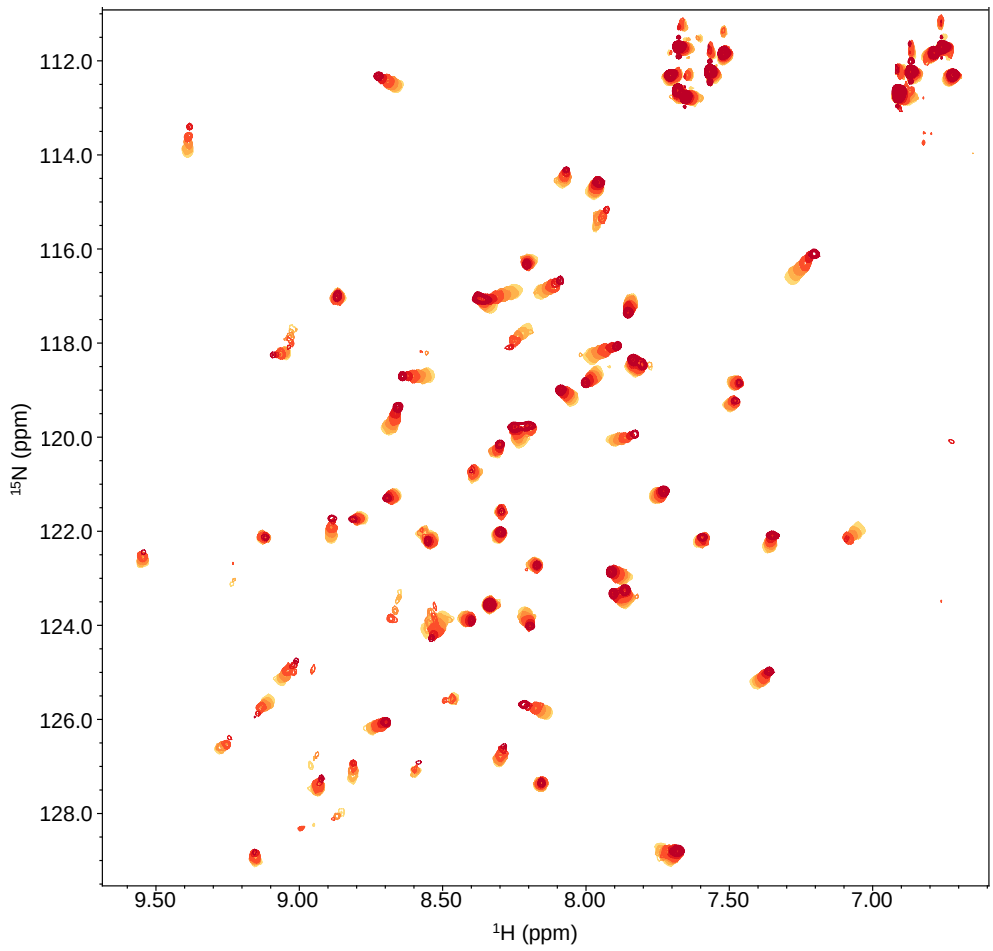


Figure 8.32: [^{1}H, ^{15}N] HSQC spectrum of a series of titration points with compound CFL02a and ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex. From light to dark shows the change in chemical shift on increasing concentration of ligand. Concentration of ligand ([L]0) ranges from 0 μM to 1.9 mM. 7 titration points were collected in total and the CSPs monitored.

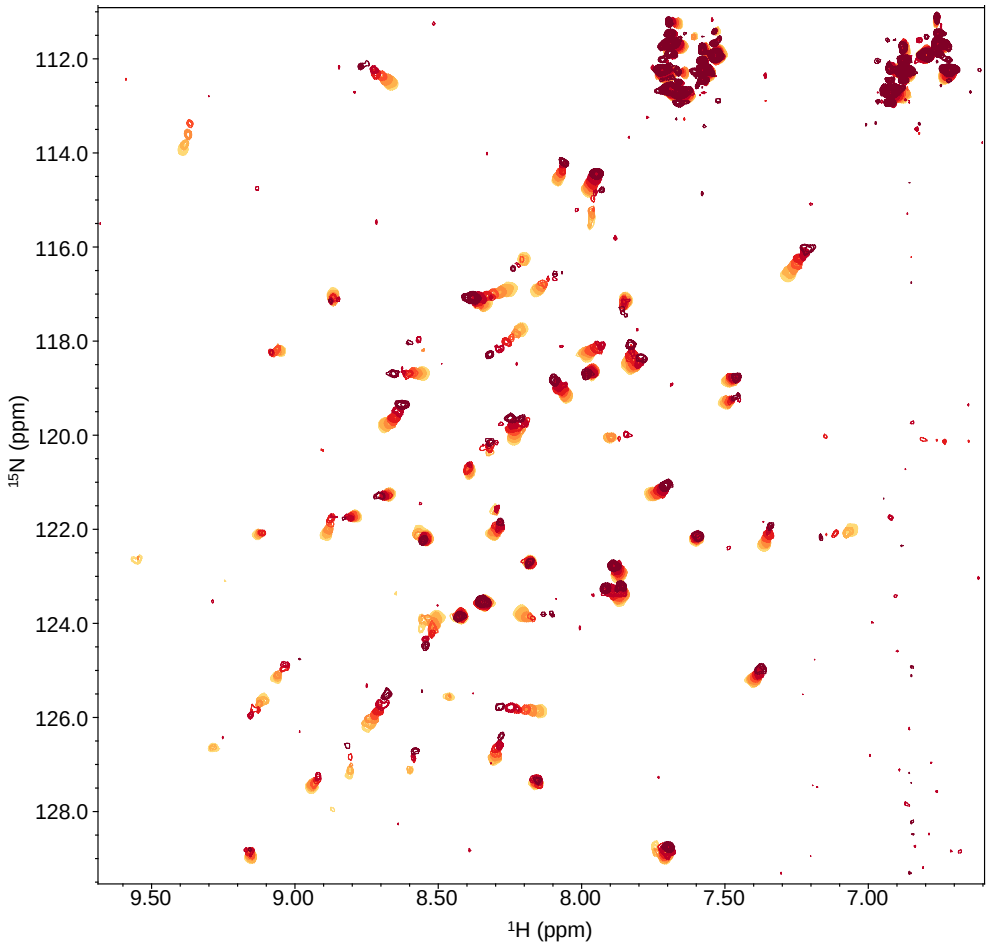


Figure 8.33: [^{1}H, ^{15}N] HSQC spectrum of a series of titration points with compound CFL02b and ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex. From light to dark shows the change in chemical shift on increasing concentration of ligand. Concentration of ligand ([L]0) ranges from 0 μM to 1.9 mM. 7 titration points were collected in total and the CSPs monitored.

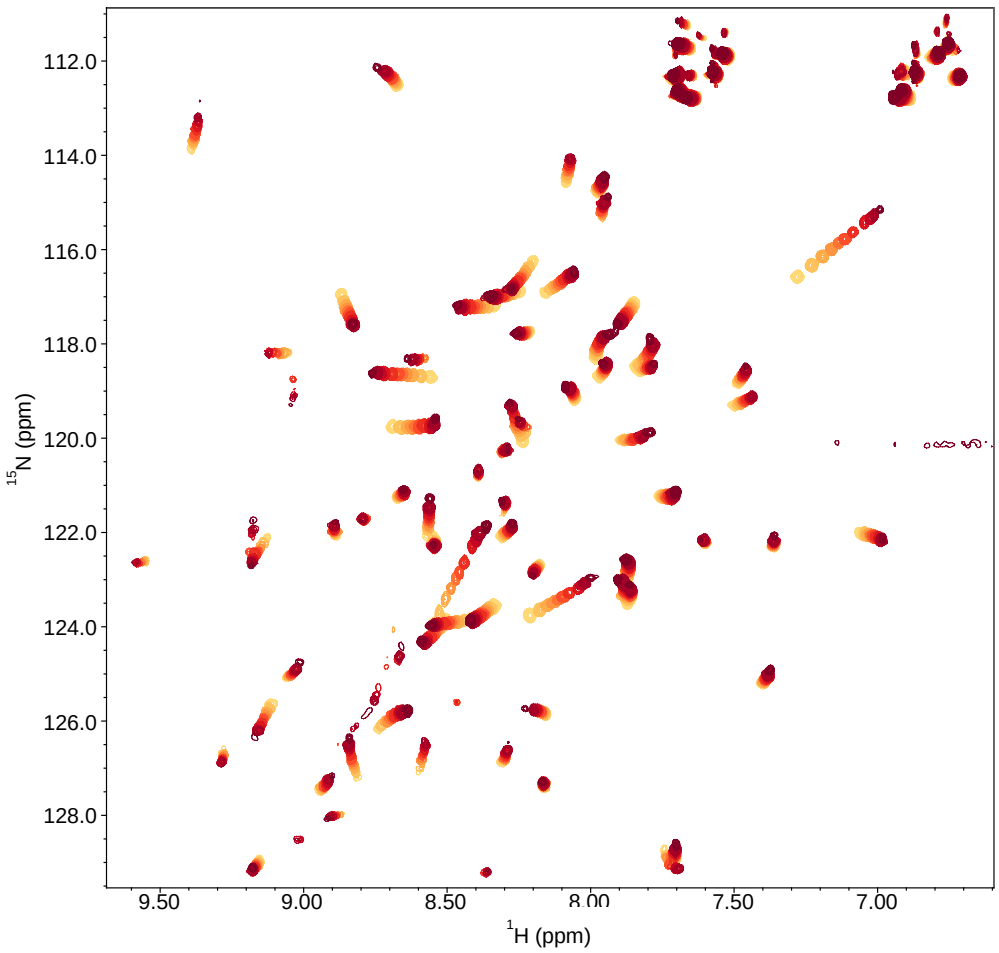


Figure 8.34: [^{1}H, ^{15}N] HSQC spectrum of a series of titration points with compound CFL02f and ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex. From light to dark shows the change in chemical shift on increasing concentration of ligand. Concentration of ligand ([L]0) ranges from 0 μM to 1.9 mM. 11 titration points were collected in total and the CSPs monitored.

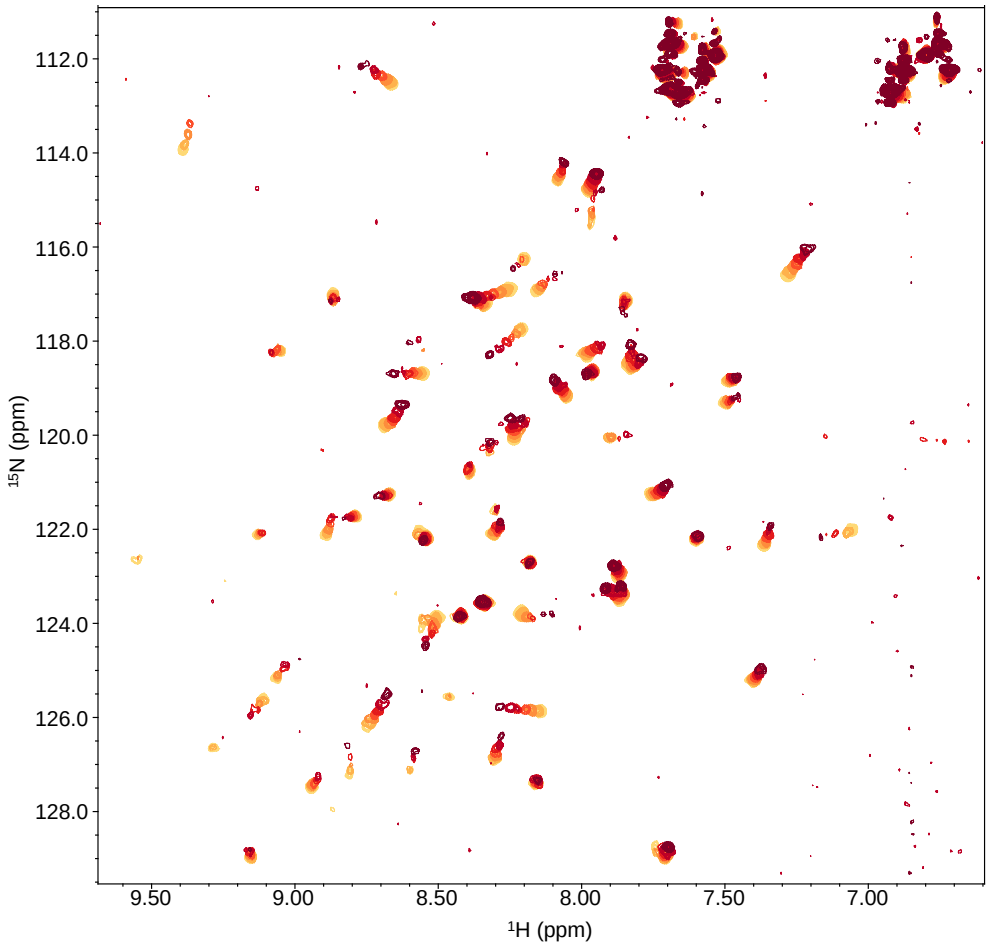


Figure 8.35: [^{1}H, ^{15}N] HSQC spectrum of a series of titration points with compound CFL17 and ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex. From light to dark shows the change in chemical shift on increasing concentration of ligand. Concentration of ligand ([L]0) ranges from 0 μM to 1.9 mM. 7 titration points were collected in total and the CSPs monitored.

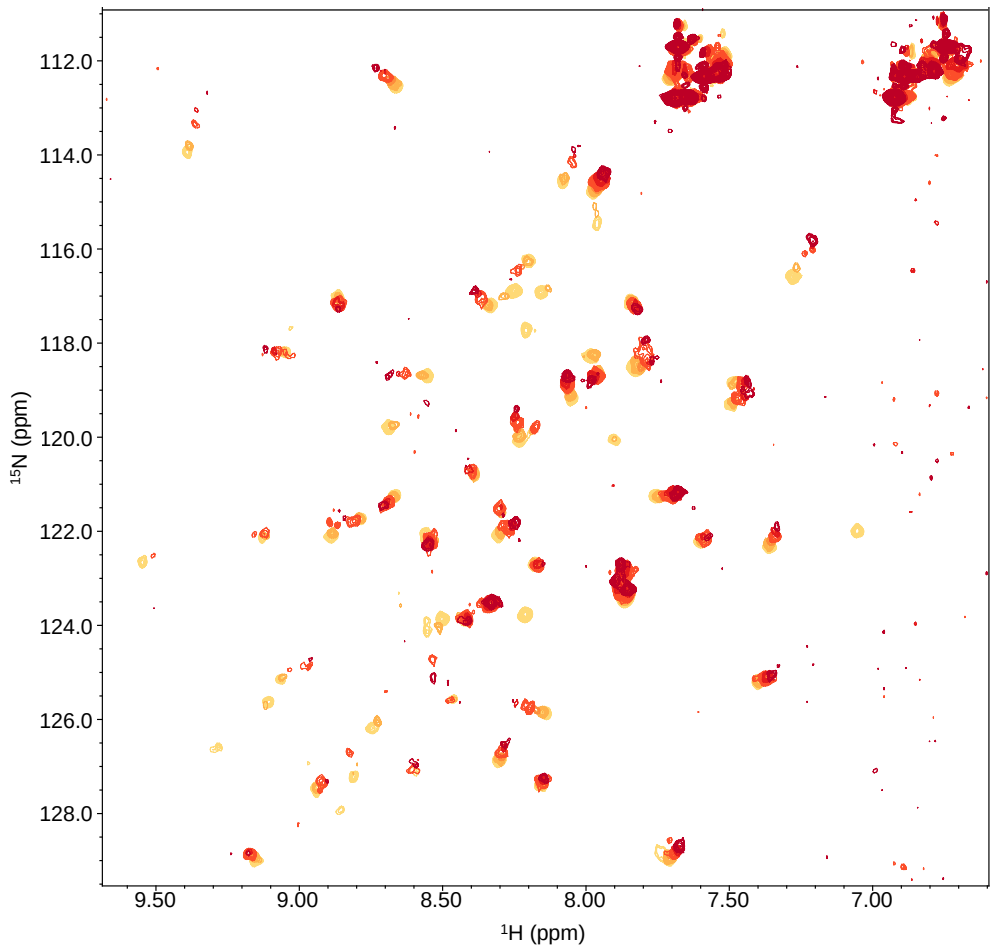


Figure 8.36: [^{1}H, ^{15}N] HSQC spectrum of a series of titration points with compound CFL17a and ^{15}N-labelled NusEΔ in the NusEΔ/NusB complex. From light to dark shows the change in chemical shift on increasing concentration of ligand. Concentration of ligand ([L]0) ranges from 0 μM to 1.9 mM. 7 titration points were collected in total and the CSPs monitored.

Chapter 9

References

1. Luch, A. *Molecular, Clinical and Environmental Toxicology: Volume 1: Molecular Toxicology*. (Birkhäuser Basel, 2009).

2. Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10,** 188 (2011).

3. Waring, M. J. *et al.* An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* **14,** 475 (2015).

4. Cohen, Y. Small Molecules: The Silent Majority of Pharmaceutical Pipelines. (2015). at <https://www.xconomy.com/boston/2015/11/23/small-molecules-the-silent-majority-of-pharmaceutical-pipelines/>

5. Dear, J. W., Antoine, D. J. & Park, B. K. Where are we now with paracetamol? *BMJ Br. Med. J.* **351,** (2015).

6. Bethesda, M. *Phase 3 Study of VX-770 Shows Marked Improvement in Lung Function Among People with Cystic Fibrosis with G551D Mutation*. (2011).

7. Wan, H. An Overall Comparison of Small Molecules and Large Biologics in ADME Testing. *Admet Dmpk* **4,** 1 (2016).

8. Fosgerau, K. & Hoffmann, T. Peptide therapeutics: current status and future directions. *Drug Discov. Today* **20,** 122–128 (2015).

9. Škalko-Basnet, N. Biologics: the role of delivery systems in improved therapy. *Biologics* **8,** 107–114 (2014).

10. Kaspar, A. A. & Reichert, J. M. Future directions for peptide therapeutics development. *Drug Discov. Today* **18,** 807–817 (2013).

11. Uhlig, T. *et al.* The emergence of peptides in the pharmaceutical business: From exploration to exploitation. *EuPA Open Proteomics* **4,** 58–69 (2014).

12. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8950–8954 (2000).

13. Walensky, L. D. *et al.* Activation of Apoptosis in Vivo by a Hydrocarbon-Stapled BH3 Helix. *Science (80-.).* **305,** 1466 LP – 1470 (2004).

14. Forum on Neuroscience and Nervous System Disorders; Board on Health Sciences Policy; Institute of Medicine. Improving and Accelerating Therapeutic Development for Nervous System Disorders: Workshop Summary. *National Academies Press (US)* (2014). at <https://www.ncbi.nlm.nih.gov/books/NBK195047/>

15. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1,** 727 (2002).

16. Hwang, H., Vreven, T., Janin, J. & Weng, Z. Protein–protein docking benchmark version 4.0. *Proteins Struct. Funct. Bioinforma.* **78,** 3111–3114 (2010).

17. Arkin, M. R., Tang, Y. & Wells, J. a. Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality. *Chem. Biol.* **21,** 1102–1114 (2014).

18. Basse, M. J. *et al.* 2P2Idb: a structural database dedicated to orthosteric modulation of protein–protein interactions. *Nucleic Acids Res.* **41,** D824–D827 (2013).

19. Major, J. Challenges and Opportunities in High Throughput Screening: Implications for New Technologies. *J. Biomol. Screen.* **3,** 13–17 (1998).

20. Barker, J., Hesterkamp, T. & Whittaker, M. Integrating HTS and fragment-based drug discovery. *Drug Discovery World* (2008). at <http://www.ddw-online.com/screening/p92852-integrating-hts-and-fragment-based-drug-discovery.html>

21. Van Dongen, M. J. P. *et al.* Structure-based screening as applied to human FABP4: A highly efficient alternative to HTS for hit generation. *J. Am. Chem. Soc.* **124,** 11874–11880 (2002).

22. Johnson, C. N., Erlanson, D. a., Murray, C. W. & Rees, D. C. Fragment-to-Lead Medicinal Chemistry Publications in 2015. *J. Med. Chem.* acs.jmedchem.6b01123 (2016). doi:10.1021/acs.jmedchem.6b01123

23. Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* **6,** 211–219 (2007).

24. Hubbard, R. E. in *Fragment-based Drug Discovery Lessons and Outlook* 1–36 (Wiley-VCH Verlag GmbH & Co. KGaA, 2016). doi:10.1002/9783527683604.ch01

25. Fink, T. & Raymond, J. L. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: Assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discove. *J. Chem. Inf. Model.* **47,** 342–353 (2007).

26. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science (80-.).* **274,** 1531–1534 (1996).

27. Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science (80-.).* **297,** 353–356 (2002).

28. Erlanson, D. a., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: The impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15,** 605–619 (2016).

29. Erlanson, D. a. & Zartler, T. Poll results: affiliation, metrics, and fragment-finding methods. *Practical Fragments Blog* (2016). at <http://practicalfragments.blogspot.co.uk/2016/10/poll-results-affiliation-metrics-and.html>

30. Gillis, E. P., Eastman, K. J., Hill, M. D., Donnelly, D. J. & Meanwell, N. a. Applications of Fluorine in Medicinal Chemistry. *J. Med. Chem.* **58,** 8315–8359 (2015).

31. Wang, J. *et al.* Fluorine in pharmaceutical industry: Fluorine-containing drugs introduced to the market in the last decade (2001-2011). *Chem. Rev.* **114,** 2432–2506 (2014).

32. Swallow, S. Fluorine in medicinal chemistry. *Prog. Med. Chem.* **54,** 65–133 (2015).

33. O'Hagan, D. Fluorine in health care: Organofluorine containing blockbuster drugs. *J. Fluor. Chem.* **131,** 1071–1081 (2010).

34. Jochen Klages, M. C. and H. K. NMR-based screening: a powerful tool in fragment-based drug discovery. *RSC Anal.* **132,** 692–705 (2006).

35. Viegas, A., Manso, J., Nobrega, F. L. & Cabrita, E. J. Saturation-transfer difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding. *J. Chem. Educ.* **88,** 990–994 (2011).

36. Mesleh, M. F. *et al.* Fragment-based discovery of DNA gyrase inhibitors targeting the ATPase subunit of GyrB. *Bioorganic Med. Chem. Lett.* **26,** 1314–1318 (2016).

37. Poornima, C. S. & Dean, P. M. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *J. Comput. Aided. Mol. Des.* **9,** 500–512 (1995).

38. Geschwindner, S. *et al.* Discovery of a novel warhead against ??-secretase through fragment-based lead generation. *J. Med. Chem.* **50,** 5903–5911 (2007).

39. Gerig, J. T. Fluorine NMR. 1–35 (2001).

40. Dalvit, C., Fagerness, P. E., Hadden, D. T. a, Sarver, R. W. & Stockman, B. J. Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.* **125,** 7696–703 (2003).

41. Gee, C. T., Koleski, E. J. & Pomerantz, W. C. K. Fragment Screening and Druggability Assessment for the CBP/p300 KIX Domain through Protein-Observed 19 F NMR Spectroscopy. *Angew. Chemie Int. Ed.* n/a–n/a (2015). doi:10.1002/anie.201411658

42. Chen, H., Viel, S., Ziarelli, F. & Peng, L. 19F NMR: a valuable tool for studying biological events. *Chem Soc Rev* **42,** 7971–7982 (2013).

43. Jehle, S., Biospin, B., Kessler, P. & Biospin, B. Fragment Screening by NMR in Drug Discovery. (2016).

44. Veronesi, M. & Dalvit, C. 19 F NMR Spectroscopy for Functional and Binding High-Throughput Screening. *Pharm. Sci.* 25–28 (2008).

45. Jordan, J. B. *et al.* Fragment Based Drug Discovery: Practical Implementation Based on 19 F NMR Spectroscopy. *J. Med. Chem.* **55,** 678–687 (2012).

46. Vulpetti, A. & Dalvit, C. Fluorine local environment: from screening to drug design. *Drug Discov. Today* **17,** 890–897 (2012).

47. Skora, L. & Jahnke, W. 19F-NMR-based dual-site reporter assay for the discovery and distinction of catalytic and allosteric kinase inhibitors. *ACS Med. Chem. Lett.* **8,** 632–635 (2017).

48. Dalvit, C. NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov. Today* **14,** 1051–1057 (2009).

49. Davies, J. W., Glick, M. & Jenkins, J. L. Streamlining lead discovery by aligning in silico and high-throughput screening. *Curr. Opin. Chem. Biol.* **10,** 343–351 (2006).

50. Winpenny, D., Clark, M. & Cawkill, D. Biased ligand quantification in drug discovery: from theory to high throughput screening to identify new biased μ opioid receptor agonists. *Br. J. Pharmacol.* **173,** 1393–1403 (2016).

51. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149–2154 (1963).

52. Geysen, H. M., Meloen, R. H. & Barteling, S. J. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Natl. Acad. Sci.* **81,** 3998 LP – 4002 (1984).

53. Lam, K. S., Lebl, M. & Krchňák, V. The 'One-Bead-One-Compound' Combinatorial Library Method. *Chem. Rev.* **97,** 411–448 (1997).

54. Lam, K. S. *et al.* A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **354,** 82 (1991).

55. Hintersteiner, M. *et al.* Single Bead Labeling Method for Combining Confocal Fluorescence On-Bead Screening and Solution Validation of Tagged One-Bead One-Compound Libraries. *Chem. Biol.* **16,** 724–735 (2009).

56. Mudd, G. Dyes, Linkers, Tags and Libraries - New Tools for Systems Chemical Biology. (University of Edinburgh, 2014).

57. Clemons, P. a. *et al.* A one-bead, one-stock solution approach to chemical genetics: Part 2. *Chem. Biol.* **8,** 1183–1195 (2001).

58. Sternson, S. M., Louca, J. B., Wong, J. C. & Schreiber, S. L. Split - pool synthesis of 1,3-dioxanes leading to arrayed stock solutions of single compounds sufficient for multiple phenotypic and protein-binding assays. *J. Am. Chem. Soc.* **123,** 1740–1747 (2001).

59. Tan, D. S., Foley, M. A., Stockwell, B. R., Shair, M. D. & Schreiber, S. L. Synthesis and Preliminary Evaluation of a Library of Polycyclic Small Molecules for Use in Chemical Genetic Assays. *J. Am. Chem. Soc.* **121,** 9073–9087 (1999).

60. Reddy, M. M., Bachhawat-Sikder, K. & Kodadek, T. Transformation of low-affinity lead compounds into high-affinity protein capture agents. *Chem. Biol.* **11,** 1127–1137 (2004).

61. Lam, K. S. *et al.* Synthesis and Screening of 'One-Bead One-Compound' Combinatorial Peptide Libraries. *Methods Enzymol.* **369,** 298–322 (2003).

62. Hintersteiner, M., Buehler, C. & Auer, M. On-bead screens sample narrower affinity ranges of protein-ligand interactions compared to equivalent solution assays. *Chemphyschem* **13,** 3472–80 (2012).

63. Hintersteiner, M. Single Beads, Single Molecules, Single Cells. (Universit of Salzburg, 2007).

64. Meisner, N.-C. *et al.* Terminal adenosyl transferase activity of posttranscriptional regulator HuR revealed by confocal on-bead screening. *J. Mol. Biol.* **386,** 435–50 (2009).

65. Hintersteiner, M. *et al.* Identification of a small molecule inhibtor of importin β mediated nuclear import by confocal on-bead screening of tagged one-bead one-compound libraries. *ACS Chem. Biol.* **5,** 967–979 (2010).

66. Hintersteiner, M. & Auer, M. A two-channel detection method for autofluorescence correction and efficient on-bead screening of one-bead one-compound combinatorial libraries using the COPAS fluorescence activated bead sorting system. *Methods Appl. Fluoresc.* **1,** 017001 (2013).

67. Hintersteiner, M. *et al.* Identification and X-ray co-crystal structure of a small-molecule activator of LFA-1-ICAM-1 binding. *Angew. Chem. Int. Ed. Engl.* **53,** 4322–6 (2014).

68. Portal, C. *et al.* Facile Synthesis of a Next Generation Safety-Catch Acid-Labile Linker, SCAL-2, Suitable for Solid-Phase Synthesis, On-Support Display and for Post-Synthesis Tagging. *Chemistryselect* **2,** 6658–6662 (2017).

69. Hintersteiner, M., Knox, A. J., Mudd, G. & Auer, M. Towards mimicking short linear peptide motifs: identification of new mixed α,β-peptidomimetic ligands for SLAM-Associated Protein (SAP) by confocal on-bead screening. *J. Chem. Biol.* **5,** 63–79 (2012).

70. Hopkins, A. L., Groom, C. R. & Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **9,** 430–431 (2004).

71. Andrews, P. R., Craik, D. J. & Martin, J. L. Functional Group Contributions to Drug-Receptor Interactions. *J. Med. Chem.* **27,** 1648–1657 (1984).

72. Kuntz, I. D., Chen, K., Sharp, K. a. & Kollman, P. a. The maximal affinity of ligands. *Proc. Natl. Acad. Sci.* **96,** 9997–10002 (1999).

73. Kucharzik, T., Lemmnitz, G., Abels, C. & Maaser, C. Tripeptide K(D)PT Is Well Tolerated in Mild-to-moderate Ulcerative Colitis. *Inflamm. Bowel Dis.* **23,** 261–271 (2017).

74. Liao, W., Chakrabarti, S., Davidge, S. T. & Wu, J. Modulatory Effects of Egg White Ovotransferrin-Derived Tripeptide IRW (Ile-Arg-Trp) on Vascular Smooth Muscle Cells against Angiotensin II Stimulation. *J. Agric. Food Chem.* **64,** 7342–7347 (2016).

75. Rajasekhar, K., Madhu, C. & Govindaraju, T. Natural Tripeptide-Based Inhibitor of Multifaceted Amyloid β Toxicity. *ACS Chem. Neurosci.* **7,** 1300–1310 (2016).

76. Reynolds, C. H., Bembenek, S. D. & Tounge, B. A. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* **17,** 4258–4261 (2007).

77. Wu, B. *et al.* High-throughput screening by Nuclear Magnetic Resonance (HTS by NMR) for the identification of PPIs antagonists. *Curr. Top. Med. Chem.* **15,** 2032–2042 (2015).

78. Richards, K. L., Rowe, M. L., Hudson, P. B., Williamson, R. A. & Howard, M. J. Combined ligand-observe 19 F and protein-observe 15 N, 1 H-HSQC NMR suggests phenylalanine as the key Δ-somatostatin residue recognized by human protein disulfide isomerase. *Sci. Rep.* **6,** 19518 (2016).

79. Daugherty, P. S. Protein engineering with bacterial display. *Curr. Opin. Struct. Biol.* **17,** 474–480 (2007).

80. Bratkovič, T. Progress in phage display: evolution of the technique and its applications. *Cell. Mol. Life Sci.* **67,** 749–767 (2010).

81. Ezkurdia, I. *et al.* Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum. Mol. Genet.* **23,** 5866–5878 (2014).

82. Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A. & Landsman, D. Histone structure and nucleosome stability. *Expert Rev. Proteomics* **2,** 719–729 (2005).

83. Conway, S. J. Bromodomains: Are readers right for epigenetic therapy? *ACS Med. Chem. Lett.* **3,** 691–694 (2012).

84. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149,** 214–31 (2012).

85. Demont, E. H. *et al.* Fragment-Based Discovery of Low-Micromolar ATAD2 Bromodomain Inhibitors. *J. Med. Chem.* **58,** 5649–5673 (2015).

86. Zou, J. X., Revenko, A. S., Li, L. B., Gemo, A. T. & Chen, H. ANCCA, an estrogen-regulated AAA ATPase coactivator for ER, is required for coregulator occupancy and chromatin modification. *PNAS* **104,** (2007).

87. Raeder, M. B. *et al.* Integrated genomic analysis of the 8q24 amplification in endometrial cancers identifies ATAD2 as essential to MYC-dependent cancers. *PLoS One* **8,** e54873 (2013).

88. Caron, C. *et al.* Functional characterization of ATAD2 as a new cancer/testis factor and a predictor of poor prognosis in breast and lung cancers. *Oncogene* **29,** 5171–81 (2010).

89. Kalashnikova, E. V *et al.* ANCCA/ATAD2 overexpression identifies breast cancer patients with poor prognosis, acting to drive proliferation and survival of triple-negative cells through control of B-Myb and EZH2. *Cancer Res.* **70,** 9402–12 (2010).

90. Revenko, A. S., Kalashnikova, E. V, Gemo, A. T., Zou, J. X. & Chen, H.-W. Chromatin loading of E2F-MLL complex by cancer-associated coregulator ANCCA via reading a specific histone mark. *Mol. Cell. Biol.* **30,** 5260–5272 (2010).

91. Ciró, M. *et al.* ATAD2 is a novel cofactor for MYC, overexpressed and amplified in aggressive tumors. *Cancer Res.* **69,** 8491–8 (2009).

92. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–73 (2010).

93. Prinjha, R. K., Witherington, J. & Lee, K. Place your BETs: the therapeutic potential of bromodomains. *Trends Pharmacol. Sci.* **33,** 146–53 (2012).

94. Brand, M. *et al.* Small Molecule Inhibitors of Bromodomain − Acetyl-lysine Interactions. (2015). doi:10.1021/cb500996u

95. Bamborough, P. *et al.* Fragment-based discovery of bromodomain inhibitors part 2: optimization of phenylisoxazole sulfonamides. *J. Med. Chem.* **55,** 587–96 (2012).

96. Vidler, L. R., Brown, N., Knapp, S. & Hoelder, S. Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *J. Med. Chem.* **55,** 7345–7359 (2012).

97. MacGurn, J. A., Hsu, P.-C. & Emr, S. D. Ubiquitin and Membrane Protein Turnover: From Cradle to Grave. *Annu. Rev. Biochem.* **81,** 231–259 (2012).

98. Matyskiela, M. E. & Martin, A. Design Principles of a Universal Protein Degradation Machine. *J. Mol. Biol.* **425,** 199–213 (2013).

99. Wenzel, D. M., Stoll, K. E. & Klevit, R. E. E2s: structurally economical and functionally replete. *Biochem. J.* **433,** 31–42 (2011).

100. Xie, C., Powell, C., Yao, M., Wu, J. & Dong, Q. Ubiquitin-conjugating enzyme E2C: a potential cancer biomarker. *Int. J. Biochem. Cell Biol.* **47,** 113–117 (2014).

101. Tokumoto, M. *et al.* Cadmium toxicity is caused by accumulation of p53 through the down-regulation of Ube2d family genes in vitro and in vivo. *J. Toxicol. Sci.* **36,** 191–200 (2011).

102. Dmitriev, A. A. *et al.* Genetic and epigenetic analysis of non-small cell lung cancer with NotI-microarrays. *Epigenetics* **7,** 502–513 (2012).

103. Molochnikov, L. *et al.* A molecular signature in blood identifies early Parkinson's disease. *Mol. Neurodegener.* **7,** 26 (2012).

104. Song, S. & Jung, Y.-K. Alzheimer's disease meets the ubiquitin–proteasome system. *Trends Mol. Med.* **10,** 565–570 (2004).

105. De Pril, R., Fischer, D. F., Roos, R. A. C. & van Leeuwen, F. W. Ubiquitin-conjugating enzyme E2-25K increases aggregate formation and cell death in polyglutamine diseases. *Mol. Cell. Neurosci.* **34,** 10–19 (2007).

106. Oh, K.-J., Kalinina, A. & Bagchi, S. Destabilization of Rb by human papillomavirus E7 is cell cycle dependent: E2-25K is involved in the proteolysis. *Virology* **396,** 118–124 (2010).

107. Lewis, M. J. *et al.* UBE2L3 Polymorphism Amplifies NF-κB Activation and Promotes Plasma Cell Development, Linking Linear Ubiquitination to Multiple Autoimmune Diseases. *Am. J. Hum. Genet.* **96,** 221–234 (2015).

108. Fiesel, F. C., Moussaud-Lamodière, E. L., Ando, M. & Springer, W. A specific subset of E2 ubiquitin-conjugating enzymes regulate Parkin activation and mitophagy differently. *J. Cell Sci.* **127,** 3488 LP – 3504 (2014).

109. Zhang, Y. *et al.* Parkin functions as an E2-dependent ubiquitin–protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. *Proc. Natl. Acad. Sci.* **97,** 13354–13359 (2000).

110. Zhao, C., Denison, C., Huibregtse, J. M., Gygi, S. & Krug, R. M. Human ISG15 conjugation targets both IFN-induced and constitutively expressed proteins functioning in diverse cellular pathways. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10200–10205 (2005).

111. Zhou, X. *et al.* Epigenetic downregulation of the ISG15–conjugating enzyme UbcH8 impairs lipolysis and correlates with poor prognosis in nasopharyngeal carcinoma. *Oncotarget* **6,** 41077 (2015).

112. Köttgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* **42,** 376 (2010).

113. Nikseresht, M. *et al.* Overexpression of the novel human gene, UBE2Q2, in breast cancer. *Cancer Genet.* **197,** 101–106 (2010).

114. Koszela, J. Novel Screening Methods for Inhibiting Ubiquitin Conjugating Enzymes. (University of Edinburgh, 2014).

115. Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* **474,** 105 (2011).

116. Huang, H. *et al.* E2 enzyme inhibition by stabilization of a low-affinity interface with ubiquitin. *Nat. Chem. Biol.* **10,** 156–63 (2014).

117. Geisler, S., Vollmer, S., Golombek, S. & Kahle, P. J. The ubiquitin-conjugating enzymes UBE2N, UBE2L3 and UBE2D2/3 are essential for Parkin-dependent mitophagy. *J. Cell Sci.* **127,** 3280 LP – 3293 (2014).

118. Sah, N. K., Khan, Z., Khan, G. J. & Bisen, P. S. Structural, functional and therapeutic biology of survivin. *Cancer Lett.* **244,** 164–171 (2006).

119. Morozumi, Y. *et al.* Atad2 is a generalist facilitator of chromatin dynamics in embryonic stem cells. *J. Mol. Cell Biol.* **8,** 349–362 (2016).

120. Sparrow, J. T. & Monera, O. D. Improvements to the TMSBr method of peptide resin deprotection and cleavage: application to large peptides. *Pept. Res.* **9,** 218–222 (1996).

121. Goldfarb, A. R., Saidel, L. J. & Mosovich, E. The ultraviolet absorption spectra of proteins. *J. Biol. Chem.* **193,** 397–404 (1951).

122. Iqbal, A. & Brown, P. J. in *Fragment-based Drug Discovery Lessons and Outlook* 355–370 (Wiley-VCH Verlag GmbH & Co. KGaA, 2016). doi:10.1002/9783527683604.ch15

123. Harner, M. J., Chauder, B. a, Phan, J. & Fesik, S. W. Fragment-based screening of the bromodomain of ATAD2. *J. Med. Chem.* (2014). doi:10.1021/jm501035j

124. Chaikuad, A., Petros, A. M., Fedorov, O., Xu, J. & Knapp, S. Structure-based approaches towards identification of fragments for the low-druggability ATAD2 bromodomain. *Medchemcomm* **00,** 1–6 (2014).

125. Bamborough, P. *et al.* Structure-Based Optimization of Naphthyridones into Potent ATAD2 Bromodomain Inhibitors. *J. Med. Chem.* 150731141956002 (2015). doi:10.1021/acs.jmedchem.5b00773

126. Chen, P. *et al.* Discovery and characterization of GSK2801, a selective chemical probe for the bromodomains BAZ2A and BAZ2B. *J. Med. Chem.* 150323104026001 (2015). doi:10.1021/acs.jmedchem.5b00209

127. Fernández-Montalván, A. E. *et al.* Isoform-Selective ATAD2 Chemical Probe with Novel Chemical Structure and Unusual Mode of Action. *ACS Chem. Biol.* **12,** 2730–2736 (2017).

128. Jeyaprakash, A. A., Basquin, C., Jayachandran, U. & Conti, E. Structural Basis for the Recognition of Phosphorylated Histone H3 by the Survivin Subunit of the Chromosomal Passenger Complex. *Structure* **19,** 1625–1634 (2011).

129. Nalepa, G., Rolfe, M. & Harper, J. W. Drug discovery in the ubiquitin–proteasome system. *Nat. Rev. Drug Discov.* **5,** 596 (2006).

130. Strickson, S. *et al.* The anti-inflammatory drug BAY 11-7082 suppresses the MyD88-dependent signalling network by targeting the ubiquitin system. *Biochem. J.* **451,** 427–437 (2013).

131. Lee, J., Rhee, M. H., Kim, E. & Cho, J. Y. BAY 11-7082 is a broad-spectrum inhibitor with anti-inflammatory activity against multiple targets. *Mediators Inflamm.* **2012,** (2012).

132. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces1. *J. Mol. Biol.* **280,** 1–9 (1998).

133. Samanta, U. & Chakrabarti, P. Assessing the role of tryptophan residues in the binding site. *Protein Eng. Des. Sel.* **14,** 7–15 (2001).

134. Doak, B. C., Norton, R. S. & Scanlon, M. J. The ways and means of fragment-based drug design. *Pharmacol. Ther.* **167,** 28–37 (2016).

135. Hann, M. M., Leach, A. R. & Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **41,** 856–864 (2001).

136. Schuffenhauer, A. *et al.* Library design for fragment based screening. *Curr. Top. Med. Chem.* **5,** 751–62 (2005).

137. Chen, I.-J. & Hubbard, R. E. Lessons for fragment library design: analysis of output from multiple screening campaigns. *J. Comput. Aided. Mol. Des.* **23,** 603–620 (2009).

138. Schulz, M. N. & Hubbard, R. E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.* **9,** 615–621 (2009).

139. Chen, I. J. & Hubbard, R. E. Lessons for fragment library design: Analysis of output from multiple screening campaigns. *J. Comput. Aided. Mol. Des.* **23,** 603–620 (2009).

140. Bohacek, R. S., McMartin, C. & Guida, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **16,** 3–50 (1996).

141. Kirkpatrick, P. & Ellis, C. Chemical space. *Nature* **432,** 823 (2004).

142. Law, R. *et al.* The multiple roles of computational chemistry in fragment-based drug design. *J. Comput. Aided. Mol. Des.* **23,** 459–473 (2009).

143. Babaoglu, K. & Shoichet, B. K. Deconstructing fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2,** 720 (2006).

144. Köster, H. *et al.* A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *J. Med. Chem.* **54,** 7784–7796 (2011).

145. Rishton, G. M. Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discov. Today* **8,** 86–96 (2003).

146. Baell, J. & Walters, M. a. Chemistry: Chemical con artists foil drug discovery. *Nature* **513,** 481–483 (2014).

147. Capuzzi, S. J., Muratov, E. N. & Tropsha, A. Phantom PAINS: Problems with the Utility of Alerts for Pan-Assay INterference CompoundS. *J. Chem. Inf. Model.* **57,** 417–427 (2017).

148. Baell, J. B. & Holloway, G. a. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53,** 2719–2740 (2010).

149. Hubbard, R. E. & Britain), R. S. of C. (Great. *Structure-based Drug Discovery: An Overview*. (Royal Society of Chemistry, 2006).

150. Lau, W. F. *et al.* Design of a multi-purpose fragment screening library using molecular complexity and orthogonal diversity metrics. *J. Comput. Aided. Mol. Des.* **25,** 621–636 (2011).

151. Enamine PPI Fragment Library. (2018). at <http://www.enamine.net/index.php?option=com_content&task=view&id=209>

152. Akritopoulou-Zanze, I. & Hajduk, P. J. Kinase-targeted libraries: The design and synthesis of novel, potent, and selective kinase inhibitors. *Drug Discov. Today* **14,** 291–297 (2009).

153. Jordan, J. B. *et al.* Fragment based drug discovery: practical implementation based on ^{19}F NMR spectroscopy. *J. Med. Chem.* **55,** 678–87 (2012).

154. Vulpetti, A., Hommel, U., Landrum, G., Lewis, R. & Dalvit, C. Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine. *J. Am. Chem. Soc.* **131,** 12949–12959 (2009).

155. Heller, S. R., McNaught, A., Pletnev, I., Stein, S. & Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J. Cheminform.* **7,** 23 (2015).

156. Khan, S. A. *et al.* Identification of structural features in chemicals associated with cancer drug response: a systematic data-driven analysis. *Bioinformatics* **30,** i497–i504 (2014).

157. VCCLAB. *Virtual Computational Chemistry Laboratory* (2005). at <http://www.vcclab.org>

158. Tetko, I. V, Tanchuk, V. Y., Kasheva, T. N. & Villa, A. E. P. Estimation of Aqueous Solubility of Chemical Compounds Using E-State Indices. *J. Chem. Inf. Comput. Sci.* **41,** 1488–1493 (2001).

159. Leroux, F. R., Manteau, B., Vors, J.-P. & Pazenok, S. Trifluoromethyl ethers – synthesis and properties of an unusual substituent. *Beilstein J. Org. Chem.* **4,** 13 (2008).

160. Yraola, F. *et al.* New efficient substrates for semicarbazide-sensitive amine oxidase/VAP-1 enzyme: Analysis by SARs and computational docking. *J. Med. Chem.* **49,** 6197–6208 (2006).

161. Rogers, D. & Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **50,** 742–754 (2010).

162. Schreyer, A. M. & Blundell, T. USRCAT: Real-time ultrafast shape recognition with pharmacophoric constraints. *J. Cheminform.* **4,** 1 (2012).

163. Black, P. E. Greedy Algorithm. *Dictionary of Algorithms and Data Structures [online]* Vreda Pieterse and Paul E. Black, eds. (2005). at <https://www.nist.gov/dads/HTML/greedyalgo.html>

164. Schulz, M. N., Landström, J., Bright, K. & Hubbard, R. E. Design of a Fragment Library that maximally represents available chemical space. *J. Comput. Aided. Mol. Des.* **25,** 611 (2011).

165. Deverell, C., Morgan, R. E. & Strange, J. H. Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame. *Mol. Phys.* **18,** 553–559 (1970).

166. Mayer, M. & Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie Int. Ed.* **38,** 1784–1788 (1999).

167. Hajduk, P. J., Olejniczak, E. T. & Fesik, S. W. One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *J. Am. Chem. Soc.* **119,** 12257–12261 (1997).

168. Dalvit, C. *et al.* Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water*. *J. Biomol. NMR* **18,** 65–68 (2000).

169. Norton, R. S., Leung, E. W. W., Chandrashekaran, I. R. & MacRaild, C. a. Applications of 19F-NMR in fragment-based drug discovery. *Molecules* **21,** (2016).

170. Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **53,** 5061–5084 (2010).

171. Böhm, H.-J. *et al.* Fluorine in Medicinal Chemistry. *ChemBioChem* **5,** 637–643 (2004).

172. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Prog. Nucl. Magn. Reson. Spectrosc.* **51,** 219–242 (2007).

173. Feng, B. Y. & Shoichet, B. K. A Detergent-Based Assay for the Detection of Promiscuous Inhibitors. *Nat. Protoc.* **1,** 550–553 (2006).

174. Edsall, J. T. & Gutfreund, H. Biothermodynamics: the Study of Biochemical Processes at Equilibrum. (1983).

175. Lepre, C. a., Moore, J. M. & Peng, J. W. Theory and applications of NMR-based screening in pharmaceutical research. *Chem. Rev.* **104,** 3641–3675 (2004).

176. Feeney, J., Batchelor, J. G., Albrand, J. P. & Roberts, G. C. K. The effects of intermediate exchange processes on the estimation of equilibrium constants by NMR. *J. Magn. Reson.* **33,** 519–529 (1979).

177. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73,** 1–16 (2013).

178. Meyer, B. & Peters, T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chem. Int. Ed.* **42,** 864–890 (2003).

179. Hajduk, P. J. *et al.* High-throughput nuclear magnetic resonance-based screening. *J. Med. Chem.* **42,** 2315–2317 (1999).

180. Granot, J. Determination of dissociation constants of 1:1 complexes from NMR data. Optimization of the experimental setup by statistical analysis of simulated experiments. *J. Magn. Reson.* **55,** 216–224 (1983).

181. Ceccarelli, D. F. *et al.* An Allosteric Inhibitor of the Human Cdc34 Ubiquitin-Conjugating Enzyme. *Cell* **145,** 1075–1087 (2011).

182. Sanders, M. a. *et al.* Novel Inhibitors of Rad6 Ubiquitin Conjugating Enzyme: Design, Synthesis, Identification, and Functional Characterization. *Mol. Cancer Ther.* **12,** 373–383 (2013).

183. Hirohama, M. *et al.* Spectomycin B1 as a Novel SUMOylation Inhibitor That Directly Binds to SUMO E2. *ACS Chem. Biol.* **8,** 2635–2642 (2013).

184. Lepre, C. a. *Practical aspects of NMR-based fragment screening*. *Methods in Enzymology* **493,** (Elsevier Inc., 2011).

185. Davies, T. G., van Montfort, R. L. M., Williams, G. & Jhoti, H. in *Fragment-based Approaches in Drug Discovery* 193–214 (Wiley-VCH Verlag GmbH & Co. KGaA, 2006). doi:10.1002/3527608761.ch10

186. Jehle, S., Kessler, P. & Biospin, B. Innovation with Integrity NMR-based Fragment Screening for Drug Discovery. (2017).

187. Dalvit, C. Ligand- and substrate-based 19F NMR screening: Principles and applications to drug discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **51,** 243–271 (2007).

188. Hajduk, P. J., Huth, J. R. & Fesik, S. W. Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. *J. Med. Chem.* **48,** 2518–2525 (2005).

189. Edfeldt, F. N. B., Folmer, R. H. A. & Breeze, A. L. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov. Today* **16,** 284–287 (2011).

190. Dalvit, C. *et al.* Sensitivity Improvement in F NMR-Based Screening Experiments : Theoretical Considerations and Experimental Applications Sensitivity Improvement in 19 F NMR-Based Screening Experiments : Theoretical Considerations and Experimental Applications. *Biochemistry* 14620–14625 (2005). doi:10.1021/ja0542385

191. Boucher, H. W. *et al.* Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **48,** 1–12 (2009).

192. Basarab, G. S. *et al.* Optimization of Pyrrolamide Topoisomerase II Inhibitors Toward Identification of an Antibacterial Clinical Candidate (AZD5099). *J. Med. Chem.* **57,** 6060–6082 (2014).

193. Skwarczynska, M. & Ottmann, C. Protein–protein interactions as drug targets. *Future Med. Chem.* **7,** 2195–2219 (2015).

194. Feng, Y., Wang, Q. & Wang, T. Drug Target Protein-Protein Interaction Networks: A Systematic Perspective. *Biomed Res. Int.* **2017,** 1289259 (2017).

195. Bush, K. *et al.* Tackling antibiotic resistance. *Nat. Rev. Microbiol.* **9,** 894 (2011).

196. Mohanty, B. *et al.* Fragment library screening identifies hits that bind to the non-catalytic surface of Pseudomonas aeruginosa DsbA1. *PLoS One* **12,** e0173436 (2017).

197. Butt, T. R. & Karathanasi, S. K. Transcription Factors as Drug Targets: Opportunities for Therapeutic Selectivity. *Gene Expr.* **4,** 319–336 (1995).

198. Ma, C., Yang, X. & Lewis, P. J. Bacterial Transcription as a Target for Antibacterial Drug Development. *Microbiol. Mol. Biol. Rev.* **80,** 139–160 (2016).

199. Drögemüller, J. *et al.* Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci. Rep.* **5,** 16428 (2015).

200. Burmann, B. M. & Rösch, P. The role of E. coli Nus-factors in transcription regulation and transcription: translation coupling: From structure to mechanism. *Transcription* **2,** 130–134 (2011).

201. Zellars, M. & Squires, C. L. Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. *Mol. Microbiol.* **32,** 1296–1304 (1999).

202. Li, S. C., Squires, C. L. & Squires, C. Antitermination of E. coli rRNA transcription is caused by a control region segment containing lambda nut-like sequences. *Cell* **38,** 851–860 (1984).

203. Nodwell, J. R. & Greenblatt, J. Recognition of boxA antiterminator RNA by the E. coli antitermination factors NusB and ribosomal protein S10. *Cell* **72,** 261–268 (1993).

204. Lüttgen, H. *et al.* Transcriptional Regulation by Antitermination. Interaction of RNA with NusB Protein and NusB/NusE Protein Complex of Escherichia coli. *J. Mol. Biol.* **316,** 875–885 (2002).

205. Greive, S. J., Lins, A. F. & Von Hippel, P. H. Assembly of an RNA-protein complex: Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in Escherichia coli. *J. Biol. Chem.* **280,** 36397–36408 (2005).

206. Robledo, R., Atkinson, B. L. & Gottesman, M. E. Escherichia coli mutations that block transcription termination by phage HK022 Nun protein. *J. Mol. Biol.* **220,** 613–619 (1991).

207. Court, D. L. *et al.* Structural and Functional Analyses of the Transcription-Translation Proteins NusB and NusE. *Microbiology* **177,** 2589–2591 (1995).

208. Luo, X. *et al.* Structural and Functional Analysis of the E. coli NusB-S10 Transcription Antitermination Complex. *Mol. Cell* **32,** 791–802 (2008).

209. Cossar, P. J. *et al.* Small-Molecule Inhibitors of the NusB-NusE Protein-Protein Interaction with Antibiotic Activity. *ACS Omega* **2,** 3839–3857 (2017).

210. Cossar, P. J. *et al.* Identification and validation of small molecule modulators of the NusB-NusE interaction. *Bioorganic Med. Chem. Lett.* **27,** 162–167 (2017).

211. Morreale, F. E. *et al.* Mind the metal: a fragment library-derived zinc impurity binds the E2 ubiquitin-conjugating enzyme Ube2T and induces structural rearrangements. *J. Med. Chem.* acs.jmedchem.7b01071 (2017). doi:10.1021/acs.jmedchem.7b01071

212. Cramer, J. *et al.* A False-Positive Screening Hit in Fragment-Based Lead Discovery: Watch out for the Red Herring. *Angew. Chemie Int. Ed.* **56,** 1908–1913 (2017).

213. Congreve, M. & Christopher, J. A. in *Fragment-based Drug Discovery Lessons and Outlook* 267–292 (Wiley-VCH Verlag GmbH & Co. KGaA, 2016). doi:10.1002/9783527683604.ch12

214. Deprez-Poulain, R. & Deprez, B. Facts, figures and trends in lead generation. *Curr. Top. Med. Chem.* **4,** 569–580 (2004).

215. Petros, A. M. *et al.* Fragment-based discovery of potent inhibitors of the anti-apoptotic MCL-1 protein. *Bioorganic Med. Chem. Lett.* **24,** 1484–1488 (2014).

216. Korczynska, M. *et al.* Docking and Linking of Fragments to Discover Jumonji Histone Demethylase Inhibitors. *J. Med. Chem.* **59,** 1580–1598 (2016).

217. Czaplewski, L. G. *et al.* Antibacterial alkoxybenzamide inhibitors of the essential bacterial cell division protein FtsZ. *Bioorganic Med. Chem. Lett.* **19,** 524–527 (2009).

218. Wolkenberg, S. E. *et al.* High concentration electrophysiology-based fragment screen: Discovery of novel acid-sensing ion channel 3 (ASIC3) inhibitors. *Bioorganic Med. Chem. Lett.* **21,** 2646–2649 (2011).

219. Petros, A. M. *et al.* Discovery of a potent and selective Bcl-2 inhibitor using SAR by NMR. *Bioorg. Med. Chem. Lett.* **20,** 6587–6591 (2010).

220. Harner, M. J., Frank, A. O. & Fesik, S. W. Fragment-Based Drug Discovery Using NMR Spectroscopy. *J. Biomol. NMR* **56,** 65–75 (2013).

221. Sledz, P. *et al.* Optimization of the Interligand Overhauser Effect for Fragment Linking: Application to Inhibitor Discovery against Mycobacterium tuberculosis Pantothenate Synthetase. *J. Am. Chem. Soc.* **132,** 4544–4545 (2010).

222. Japan, T. N. M. R. S. *Experimental Approaches of NMR Spectroscopy: Methodology and Application to Life Science and Materials Science*. (Springer Singapore, 2017).

223. Williamson, M. P. *Chapter 3 Applications of the NOE in Molecular Biology*. *Annual Reports on NMR Spectroscopy* **65,** (Elsevier Ltd, 2009).

224. Strauss, M. J. The Nitroaromatic Group in Drug Design. Pharmacology and Toxicology (for Nonpharmacologists). *Ind. Eng. Chem. Prod. Res. Dev.* **18,** 158–166 (1979).

225. Kehrenberg, C., Schwarz, S., Jacobsen, L., Hansen, L. H. & Vester, B. A new mechanism for chloramphenicol, florfenicol and clindamycin resistance: methylation of 23S ribosomal RNA at A2503. *Mol. Microbiol.* **57,** 1064–1073 (2005).

226. Zeng, L. *et al.* Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. *J. Am. Chem. Soc.* **127,** 2376–7 (2005).

227. Pasternak, A. Cambridge MedChem Consulting. (2012). at <https://www.cambridgemedchemconsulting.com/resources/bioisoteres/>

228. Wirth, M., Zoete, V., Michielin, O. & Sauer, W. H. B. SwissBioisostere: a database of molecular replacements for ligand design. *Nucleic Acids Res.* **41,** D1137–D1143 (2013).

229. Weidemann, T., Seifert, J.-M., Hintersteiner, M. & Auer, M. Analysis of protein-small molecule interactions by microscale equilibrium dialysis and its application as a secondary confirmation method for on-bead screening. *J. Comb. Chem.* **12,** 647–54 (2010).

230. Madigan, M. T. & Martinko, J. M. *Brock Biology of Microorganisms*. (Pearson Prentic Hall, Inc., 2006).

231. Al-Ghorbani, M., Bushra Begum, a., Zabiulla, Z., Mamatha, S. V. & Khanum, S. A. Piperazine and morpholine: Synthetic preview and pharmaceutical applications. *Res. J. Pharm. Technol.* **8,** 611–628 (2015).

232. Hroch, L. *et al.* Benzothiazoles - Scaffold of Interest for CNS Targeted Drugs. *Current Medicinal Chemistry* **22,** 730–747 (2015).

233. Kamal, A., Syed, M. A. H. & Mohammed, S. M. Therapeutic potential of benzothiazoles: a patent review (2010 – 2014). *Expert Opin. Ther. Pat.* **25,** 335–349 (2015).

234. Singh, M. K., Tilak, R., Nath, G., Awasthi, S. K. & Agarwal, A. Design, synthesis and antimicrobial activity of novel benzothiazole analogs. *Eur. J. Med. Chem.* **63,** 635–644 (2013).

235. Doble, A. The pharmacology and mechanism of action of riluzole. *Neurology* **47,** 233S–241S (1996).

236. Trevor, A. J. & Katzung, B. G. *Pharmacology examination & board review*.

237. Droegemueller, Jo., Schweimer, K. & Roesch, P. Solution Structure of NusE (S10) from Thermotoga maritima. (2013). doi:10.13018/BMR19533

238. Schumann, F. H. *et al.* Combined chemical shift changes and amino acid specific chemical shift mapping of protein–protein interactions. *J. Biomol. NMR* **39,** 275–289 (2007).

239. Zuiderweg, E. R. P. Mapping Protein−Protein Interactions in Solution by NMR Spectroscopy. *Biochemistry* **41,** 1–7 (2002).

240. Burmann, B. M. *et al.* A NusE:NusG Complex Links Transcription and Translation. *Science (80-.).* **328,** 501 LP – 504 (2010).

241. Van Zundert, G. C. P. *et al.* The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428,** 720–725 (2016).

242. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: a protein− protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125,** 1731–1737 (2003).

243. Rolland, T. *et al.* A Proteome-Scale Map of the Human Interactome Network. *Cell* **159,** 1212–1226 (2014).

244. Murray, C. W. & Blundell, T. L. Structural biology in fragment-based drug design. *Curr. Opin. Struct. Biol.* **20,** 497–507 (2010).

245. Wang, Z.-X. An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Lett.* **360,** 111–114 (1995).

246. Glaser, F. *et al.* ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19,** 163–164 (2003).

247. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33,** W299–W302 (2005).

248. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529–W533 (2010).

249. Celniker, G. *et al.* ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Isr. J. Chem.* **53,** 199–206 (2013).

250. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44,** W344–W350 (2016).

251. Higueruelo, A. P. *et al.* Atomic interactions and profile of small molecules disrupting protein–protein interfaces: the TIMBAL database. *Chem. Biol. Drug Des.* **74,** 457–467 (2009).

252. Cayot, P. & Tainturier, G. The Quantification of Protein Amino Groups by the Trinitrobenzenesulfonic Acid Method: A Reexamination. *Anal. Biochem.* **249,** 184–200 (1997).

253. Altamore, T. M. *et al.* Random-Coil:α-Helix Equilibria as a Reporter for the LewisX–LewisX Interaction. *Angew. Chemie Int. Ed.* **50,** 11167–11171 (2011).

254. Johnson, B. A. in *Protein NMR Techniques* 313–352 (Springer, 2004).